Does Social Similarity Pay Off? Homophily and Venture Capitalists' Deal Valuation, Downside Risk Protection and Financial Returns in India

Kim Claes
SKK Graduate School of Business
Sungkyunkwan University
03063 Seoul, Korea
Ph: + 82 (0)2 740 1841
E-Mail: kim.claes@skku.edu

Balagopal Vissa
INSEAD
1 Ayer Rajah Avenue
Singapore 138676
Tel: (+65) 6799 5388
E-mail: balagopal.vissa@insead.edu

Accepted for Publication in *Organization Science*

Keywords: Entrepreneurship, Venture Capital, Interorganizational Relationships, Non-parametric Statistics, Social Similarity, Caste

July 2019

We thank Henrich Greve, Phanish Puranam, participants at the INSEAD Doriot Entrepreneurship Conference, Academy of Management meetings, and seminar participants at the Lee Kong Chian School of Business (Singapore Management University), ESSEC Business School, Imperial College Business School and Judge Business School (University of Cambridge) for valuable feedback on earlier versions of this manuscript. We also thank Lily Fang for helpful advice. Remaining errors are our own. A previous version of this manuscript was a finalist for the OMT Division's Best Entrepreneurship Paper Award, at the Academy of Management's annual meeting in Atlanta in August 2017. The second author gratefully acknowledges financial support from INSEAD's Emerging Markets Institute.

Does Social Similarity Pay Off? Homophily and Venture Capitalists' Deal Valuation, Downside Risk Protection and Financial Returns in India

We ask how social similarity between start-up founders and venture capitalists (VCs) influences VCs' pricing decisions and returns on investments. We conceptualize how regional and caste similarity, two salient aspects of social similarity in India, affect two distinct aspects of deal pricing: pre-money valuation and investors' downside risk protection in the Indian venture capital market. We theorize that VCs reflect the benefits and costs of social similarity by setting higher pre-money valuation when investing in companies led by socially similar founders while also minimizing their downside risks in these investments. We expect social similarity's impact on pricing is amplified when VCs face greater subjective uncertainty, such as for early-stage deals or if the VCs lack expertise in the start-up company's product market. Finally, we claim that VCs achieve superior returns on investments when their deal pricing accurately reflects the impact of social similarity. We tested our conceptual model using both parametric and nonparametric methods on a hand-collected dataset of all deals that occurred during 2005 to 2012, and we supplemented our analyses with in-depth, qualitative interviews that contextualize our findings. The pattern of findings on regional similarity are consistent with our model, but the effects of caste in our data are theoretically anomalous. Post hoc analyses to resolve the anomaly suggest an "intrinsic quality" mechanism, whereby higher-caste VCs set higher valuations when matching with lower-caste founders that signal high quality. Overall, our findings offer evidence that VCs incorporate social attributes into deal pricing in nuanced yet boundedly rational ways.

The financing of start-up companies is fundamental to the entrepreneurial process. Cash-constrained entrepreneurs strive to secure financial resources from investors in a context of significant uncertainty (Gompers and Lerner 2001). Venture capital firms' financial returns on their investments depend on the ability of venture capitalists (VCs)[1] to identify the right start-up companies to fund, price the deals appropriately, and subsequently monitor and mentor these portfolio companies. Many studies of how VCs select a start-up company (henceforth 'company') that needs financing have shown that VCs are influenced by attributes of the company such as the attractiveness of its product market (Sorenson and Stuart 2001), its technological capabilities (Hsu and Ziedonis 2013) as well as the founders' human and social capital (Hallen 2008) and networking abilities (Hallen 2008; Hallen and Eisenhardt 2012; Vissa 2012). While management scholars have explored how VCs select companies to fund and the performance consequences for the focal company (Gulati and Higgins 2003; Pahnke et al. 2014), management scholarship has paid less attention to equally important pieces of the puzzle: What determines the price VCs charge for the capital they provide and what financial returns do VCs earn on these capital investments?

Because of the theoretical peculiarities underlying VC markets, answering these questions on deal pricing and returns is critical. On the one hand, the venture capital market is a financial capital market in which the flow of capital allocated at the right price and to its best use is considered the norm (Gompers et al. 2008)). Some research provides evidence for this principle because studies show that the features of companies such as their market position and capabilities are significant drivers of valuation (Ge et al. 2005). Yet, unlike public financial markets, the market for venture capital is a thin and illiquid private capital market (Hochberg et al. 2010) characterized by high uncertainty and no reliable, 'right' market price for the securities issued by a company. Furthermore, building a valuable company requires a strong collaborative relationship between company founders and VCs (Cable and Shane 1997; Huang and Knight 2017). These factors suggest that the actors' social attributes may play a significant role in deal pricing and returns.

Initial attempts to explore the economic consequences of social similarity between company founders and VCs have yielded somewhat inconsistent findings. In particular, scholars have focused on how shared ethnicity

[1] We use the term 'venture capitalists' or 'VCs' to refer to the general partners of venture capital firms.

influences VCs' selection of companies for capital injection, how they price that capital and the VCs' performance outcomes. However, the results of this research – using data from the US – are sometimes conflicting. Hegde and Tumlinson (2014) suggest that VCs rationally select co-ethnic-led companies for funding and mentor them better, yielding superior returns to the VCs. In contrast, Bengtsson and Hsu (2015) argue that non-rational decision processes may motivate VCs to invest in co-ethnic-led companies. They obtain empirical evidence that VCs systematically select co-ethnic-led companies and pay them more (i.e. provide higher valuations) yet experience poorer performance from such investments. Moreover, Zhang et al. (2016) posit that both ethnic similarity and ethnic *dissimilarity* influence pricing. Their research offers evidence that ethnic-minority (Asian-led) VCs pay significantly more (i.e., provide higher valuations) to invest into ethnic-majority (white), founder-led companies, although they accrue no direct performance benefits from doing so. In summary, it is unclear whether and how social attributes of VCs and company founders matter for important venture capital market outcomes, such as deal pricing and financial performance.

We investigate how social attributes influence economic outcomes in VC markets by expanding the conversation in two ways. First, we suggest that deal pricing in VC markets goes beyond valuation to include contractual terms, such as liquidation preferences, cumulative dividend rights, redemption rights, and anti-dilution rights, that regulate the VC's downside risk protection. Thus, we examine the effects of social similarity on both the valuation aspect as well as the downside risk protection aspect of pricing. Second, situating our empirical study in the world's third largest entrepreneurial ecosystem, India, enables us to examine two important and understudied social attributes in the Indian setting: individuals' regions of origin and their castes. Our conceptual framework applies well-established mechanisms of homophily to develop theoretical predictions on the effect of our novel social similarity attributes (region and caste) on valuation and downside risk protection in Indian VC deals as well as on subsequent financial returns to VCs. In post hoc analyses, we delve deeper into theoretically anomalous findings in our data pertaining to the role of caste in VC markets.

Given that India is a less studied empirical context, we grounded our model through in-depth qualitative interviews with six VCs and founders in India. Our quantitative findings utilize a unique, hand-collected dataset of 622 VC deals in Indian companies between 2005 and 2012, a period when a change in regulations required regulatory filings that in turn enabled us to accurately compute deal valuations, downside risk protection (Bengtsson and Sensoy 2011), and VCs' internal rates of return (IRR). Our analyses include both parametric and nonparametric approaches. The parametric analyses utilize a Heckman-type selection model to account for the VCs' determinants of company selection and subsequently examine how regional and caste similarity between VCs and company founders affect valuation, downside risk protection as well as the VCs' IRR on these investments. In supplementary, nonparametric analysis we use two-sided matching models with Bayesian (MCMC) estimation (Sørensen 2007) to investigate the effect of social similarity on IRR after accounting for the initial two-sided matching of companies and VCs based on both observed and unobserved features.

Our nuanced exploration of the economic impact of actors' social attributes offers three contributions to the strategy and organizations literature. First, we shed light on the drivers of the crucial yet understudied outcome of deal pricing in venture capital markets. We explore how social similarity affects both the valuation as well as downside risk protection aspects of deal pricing and the subsequent financial returns VCs earn on these deals. Second, we obtain robust evidence that the well-established mechanisms of homophily on *nominal* social attributes also apply to the Indian context where the relevant social attribute is regional similarity. Our results here paint a clear picture of boundedly rational VCs looking ahead to anticipate the benefits and costs of regional similarity in building a collaborative partnership and reflecting these benefits and costs in the deal pricing in economically significant ways. Finally, our results present initial tentative evidence of the complex effects of caste similarity on VC pricing. Our findings on caste paint a blurry picture for two reasons. First, the statistical significance of caste as a driver of pricing and returns is much lower than that of regional similarity in our data. This suggests that, at least in the technology entrepreneurship domain of the Indian economy, an individual's caste is perhaps a less important social attribute than the Indian region that individuals come from. Second, caste effects, when caste is treated as a rank-ordered attribute, are complex but still seem more consistent with boundedly rational behavior by VCs than with non-rational bias.

THEORY AND HYPOTHESES

Pricing and Performance Outcomes in Venture Capital Markets

Investments in new companies occur in discrete financing rounds, in which a VC firm (or syndicate of VC firms) injects capital into a selected company in exchange for an ownership stake. Raising money in discrete rounds is an important governance tool that ties the VCs' capital injection to measurable value-creation milestones of the company, while providing opportunities to manage value appropriation among shareholders (VCs and founders) via the allocation of property rights to the company's earnings (Kaplan and Strömberg 2003). Such investments yield a financial return (or a loss) to VCs when the company experiences a liquidity event: either a successful initial public offering (IPO), a successful trade sale, a failure-induced liquidation of assets, or a distress sale. The VCs' return on investment is a function of their initial choice of company and deal pricing as well as the subsequent collaboration with founders to build a thriving business (Cable and Shane 1997; Sørensen 2007).

A voluminous management literature explores the processes by which VCs identify promising companies to fund (cf. Drover et al. 2017 for a comprehensive review). In contrast, we know much less about how VCs set the prices and terms of their investment or the VCs' subsequent financial performance, the key outcomes we examine. Because of the enormous uncertainty around early-stage companies, pricing of their investment capital requires considerable business acumen from VCs. As we elaborate below, scholars have focused on two distinct aspects of pricing: the valuation at which VCs inject capital (Bengtsson and Hsu 2015; Zhang et al. 2016), and the contractual terms underpinning the VCs' downside risk protection in the event of poor company performance (Bengtsson and Sensoy 2011).

The Valuation Aspect of Pricing

Valuation is an important characteristic of the price charged for capital allocation in VC markets. The dollar amount of an investment in the company divided by the equity stake gained in that investment is known as post-money valuation. *Pre-money valuation* is the post-money valuation minus the investment amount. A higher valuation (either pre- or post-money) implies a higher price from the VC's point of view since the equity stake is lower for a given investment amount.

The relatively sparse research on drivers of valuation is based on US data and suggests that industry structure and resources – of both VC firms and companies – determine valuation. Working from the supply side, scholars have found that greater availability of capital from limited partners (Gompers and Lerner 2000) and greater competitive rivalry caused by the entry of new VC firms (Hochberg et al. 2010) drive up valuations. However, VC firms enjoying higher status obtain lower valuations (Ge et al. 2005; Hsu 2004). Furthermore, research on the demand side suggests that industry structure and the internal resources and capabilities of the focal company (Ge et al. 2005; Hsu 2004) are also significant valuation drivers.

The Downside Risk Protection Aspect of Pricing

Another crucial feature of pricing in VC markets is the degree to which VC firms can *ex ante* design contracts to align incentives with founders and protect themselves from an important endogenous source of uncertainty: potential agency conflicts between VCs and founders. Drawing on contract theory (Hart 2001; Hölmstrom 1979), finance scholars (Kaplan and Strömberg 2003; Sahlman 1990) have documented how features of VC contracts in the US, such as the allocation of cash flow rights, liquidation rights, and control rights, mitigate agency conflicts between VCs and founders.

Ex ante contract design enables VCs to regulate their ownership stakes in a company (and thus the price they charge for their capital) by linking the company's future performance to agreed milestones. Indeed, Kaplan and Stromberg (2003) report that VC ownership in a company could increase by as much as 8.8% on average from a neutral-to-negative company performance event. VCs endeavor to limit downsides to their cash flow rights in the event of poor company performance through contractual terms such as liquidation preference rights, redemption rights, cumulative dividend rights, etc., that effectively increase their ownership stake in a manner contingent on future company performance. Bengtsson and Sensoy (2011) combine these individual contractual terms to conceptualize *downside risk protection* as the extent to which VCs possess a more debt-like claim on the company's cash flows. In essence, the concept of downside risk protection captures the likely degree of dispersion in VCs ownership stakes (and thus the price of their capital) based on company performance. This conception thus sheds light on the extent to which VCs' cash flow rights are protected beyond the ownership stake implied by a particular pre-money valuation.

VCs' Financial Performance

Scholars (Bengtsson and Hsu 2015; Sørensen 2007) acknowledge that VCs' financial returns in a specific investment are influenced by their choice of company, the pricing of the investment and the subsequent collaboration with company founders to build a thriving business. Sørenson's (2007) study offers persuasive evidence that both initial matching and subsequent collaboration are statistically and economically significant drivers of VCs' financial success, as proxied by successful IPO or trade sales of their portfolio companies. However, it is unclear whether the beneficial effects of matching are a result of VCs setting favorable pricing (lower valuation and better contractual terms) or the effects of acquiring investment access to more promising companies, even if the initial pricing and terms are less favorable to the VCs. Indeed, although Kerr et al. (2014) imply that valuations in the initial financing rounds are less important to VCs' eventual financial performance, others suggest that both valuation and downside risk protection likely have a significant effect on VCs' performance (Gornall and Strebulaev 2017). Determining how social similarity affects VCs' investment performance is theoretically important to understanding how the initial selection of promising companies and subsequent ongoing collaboration with founders are related to VC performance in a specific investment.[2]

Social Similarity as a Driver of Pricing and VC Performance Outcomes

Three prior studies, Hegde and Tumlinson (2014), Bengtsson and Hsu (2015), and Zhang et al. (2016), examine social similarity's effects on VC's investments and performance. These works consider a specific type of similarity – co-ethnicity – between VCs and founders in the United States. Taken together, their results offer a somewhat inconsistent account of the influence of social similarity on VCs' deal pricing and performance. As we

[2] Some research focuses on VCs' financial performance based on returns at the *portfolio* level. But given our research question and in line with Sørenson (2007) and other studies examining the role of social attributes on VC outcomes, we examine returns at the level of investments into specific companies, i.e., at the VC-company dyad level of analysis.

will show, these inconsistencies stem from different conceptualizations of ethnicity, divergent theoretical mechanisms linking ethnicity to pricing and performance, and differences in samples and measurements of the same concept.

Ethnicity as a Nominal Attribute and the Downside of Favoring Co-Ethnics

Bengtsson and Hsu (2015) examine ethnicity effects by treating it as a nominal category, i.e., their measures assume no rank ordering among ethnicities, such as white American, African American, Hispanic American, etc. The mechanism linking co-ethnicity to VCs' pricing and performance is considered an unconscious social bias. Their central argument assumes that VCs hold economic power over founders seeking funding and suggests that VCs irrationally pay more for deals (set higher valuations) and set founder-friendlier terms (less downside risk protection for VCs) when the founder is a co-ethnic. As a result, VCs experience poorer results which is indicated by co-ethnicity's strong negative effect on the likelihood of exits. In essence, Bengtsson and Hsu (2015) evoke two mechanisms. First, VCs may fall prey to nonrational processes whereby actors are (unconsciously) generous to others like themselves in bargaining situations (Homans 1950). Second, VCs may fail to effectively monitor co-ethnic-led companies because homophily (McPherson et al. 2001) induces VCs to overestimate the capabilities of co-ethnic entrepreneurs, turn a blind eye, or misattribute poor performance to bad luck rather than lack of effort or competence. In effect, the costs of co-ethnicity overwhelm any putative benefits. Bengtsson and Hsu (2015: 352) observe: "These empirical patterns suggest not only that there is a negative average association between co-ethnics and exit likelihood but also that we are unable to find a positive relationship when we condition the sample based on VC or company characteristics…our above explanation is of irrational overinvesting by co-ethnics and lack of monitoring (direct governance and/or allocating more favorable financial contract terms to the entrepreneur)".

Ethnicity as a Nominal Attribute and the Benefits of Favoring Co-Ethnics

Hegde and Tumlinson (2014) also view ethnicity as a nominal category and implicitly assume VCs hold economic power over the founders but they suggest a different mechanism. They claim that rational reasons explain VCs' superior performance outcomes when investing in co-ethnic entrepreneurs. According to their theory co-ethnicity gives VCs a superior ability to identify noisy signals of quality, thus separating 'the wheat from the chaff" when selecting entrepreneurs. In addition, VCs collaborate more effectively with co-ethnic entrepreneurs thanks to homophily's beneficial effects of easier communication and accelerated trust formation (McPherson et al. 2001). Consistent with this mechanism, their study provides empirical evidence that VCs experience more successful exits when investing in co-ethnic-led companies. Since they do not consider co-ethnicity's effects on pricing, we do not know whether VCs ex ante set higher valuations or more onerous contractual terms that mitigate downside risks when investing into co-ethnic-led companies, although that is the logical implication. In short, the benefits of co-ethnicity outweigh any putative costs in this view.

Ethnicity as a Rank-Ordered Attribute and Favoring Ethnic Dissimilarity as a Functional Response

While Zhang et al. (2016) also assume that VCs hold economic power over founders, they conceptualize ethnicity as a rank-ordered social attribute in the US with Asian-ethnic-led VCs having lower social status than mainstream (white-ethnic-led) VCs. Using this rank-ordered framework, Zhang et al. (2016) argue that Asian VCs set higher valuations when investing in white-ethnic-led companies. They reason that because access to deal flow is stratified by social status, Asian VCs are motivated to fund white-ethnic-led companies as a way to improve their own lower social status (and thus access better deals in the future). Hence, they offer higher pre-money valuation as an incentive for white-ethnic-led companies to transact with Asian-ethnic-led VCs[3].

Taken together, this body of work examines the effect of a specific social attribute, ethnicity, because of its significance in the United States. However, the appropriate bases of social similarity depend on the specific context. In this paper we propose that the salient aspects of social similarity in the Indian context are region and caste, and we conceptualize these attributes as nominal (non-rank-ordered) social categories.

[3] Zhang et al (2016) uses a sample of only Silicon Valley VC firms; Bengtsson and Hsu (2015) use a sample of all US funds and companies, whilst Hegde and Tumlinson (2014) include both foreign and US funds and companies. In addition, the three studies differ in the granularity of ethnicity. Hegde and Tumlinson (2014) use a definition that includes a larger number of ethnicities while the other two use a tighter definition with fewer ethnicity categories. These sampling frame and measurement choices make it harder to cumulate the findings across the three studies.

Social Similarity in India: Region and Caste

Contemporary India is highly diverse and founded upon an ancient civilization. Superimposed upon traditions and customs going back more than 5,000 years (Keay 2011), the modern Indian republic since 1950 is a Western-style democracy and a market economy since 1991 (Sunil 1997). Although India has a fast-growing entrepreneurial ecosystem, the venture capital industry is young (dating back to 2000) and has limited role models or institutionalized practices. Given the blend of the modern VC market and ancient civilization, India's entrepreneurial ecosystem has a distinct social structure (relative to the more widely studied Western context) where the relevant bases of social similarity are region and caste (Vissa 2011).

Regional similarity: India's social structure is characterized by dramatic regional heterogeneity. This variation in regional culture is manifested through the sheer variety of its languages. In India 18 different languages are officially recognized, 12 of which are spoken by over ten million people. Because Indian regions were reorganized along linguistic lines following independence from the British, regional language is strongly related to regional culture and customs (Ramaswamy 1997). Regional language diversity largely explains why English is often the *lingua franca* when different groups communicate (Kachru 1983). In contemporary India an individual's knowledge and use of a regional language is a strong signal of regional identity (Krishnamurti 1990). That is, actors drawn from the same region (as proxied by their regional language) are likely to share the same taken-for-granted cultural assumptions and values. Thus, regions serve as cultural markers that generate solidarity and cohesion between individuals. Since there is no pecking order among regions in this setting, regional similarity is more properly understood as a nominal attribute.

Caste similarity: One of the most salient institutional features of Indian society is caste. Historically, the caste system's primary function was to regulate interpersonal interactions in religious, social and economic life by creating a hierarchical division of society into five endogamous groups (termed *varnas*) and specifying permissible interactions across them (Ghurye 1969; Srinivas 1957). At the top of the hierarchy are priests (*brahmins*), warriors (*kshatriyas*) and merchants (*vaishyas*). These groups historically enjoyed significant material resources and freedoms; today they are termed 'forward castes' by the Indian government and do not possess any special privileges in the Indian state's affirmative action programs. Other castes, the peasants (*shudras*) and the untouchables (now called *dalits* and formally outside of the caste system), historically faced significant discrimination, particularly the *dalits*. These groups, respectively designated 'backward caste' and 'scheduled caste', are beneficiaries of the government's affirmative action programs.

Scholars who study caste in historical Indian society either view it as a status hierarchy (Dumont 1980) based on ideological rather than material resources, or as a social system for economic exploitation (Wiser 1988). However, the caste system as practiced in contemporary Indian society is considered a way of demarcating non-hierarchical, differentiated groups (*jatis*) that exhibit local dominance in economic or political terms, irrespective of their formal hierarchical position in the normative *varna* scheme (Srinivas 1957). Like prior management research (Chen et al. 2015; Damaraju and Makhija 2018; Vissa 2011) we adopt this latter view of caste, conceptualizing it as a nominal attribute consisting of the following five distinctive caste groups: priests, warriors, merchants, peasants and untouchables.

Social Similarity and Pricing in Indian Venture Capital Markets

Consistent with prior research, we assume that financial capital is a scarce resource and hence a binding constraint, and that VCs as providers of capital wield economic power over resource-starved entrepreneurs seeking capital. Indeed, our qualitative interviews suggest that during the period of our study (2005-2012) VCs were by far the stronger party in the economic exchange because founders rarely had more than a single term sheet on the table and likely were faced with a "take it or leave it" offer.[4] We, therefore, develop our arguments from the perspective of VCs as the final arbiters of observed prices in this two-sided matching market.

<u>Main Effects of Regional and Caste Similarity on Pricing</u>

Research on homophily has reported that the key benefits of social similarity are accelerated trust formation (Brewer 1979) and better communication (Rogers and Bhowmik 1970; Zucker 1986), which increase

[4] This feature enables econometric identification and is valid for our study period. Greater availability of venture capital in India has increased the number of founding teams receiving multiple simultaneous term sheets in recent years. The asymmetry of economic power between VCs and founders is a boundary condition of our theorizing.

the odds of creating an effective working relationship (Gabarro 1990). In our context, these propositions suggest that Indian VCs, as boundedly rational actors, treat regional similarity or caste similarity as a noisy signal of the odds of successfully creating an effective working relationship with founders. A productive working relationship enables VCs and founders to collaborate on the highly uncertain task of building a successful venture without falling prey to dysfunctional conflict (Forbes et al. 2010), thereby generating superior venture performance outcomes. There is evidence of the influence of social similarity on economic action in our empirical context. Vissa (2011) reports Indian start-up founders form economic exchange ties with alliance partners and customers on the basis of regional similarity. Likewise, he finds Indian start-up founders have intentions to connect with potential customers and alliance partners based on caste similarity. These findings suggest that social similarity might be important for matching with investors as well. Indeed, our qualitative interviews suggest a link between judgments about ease of collaboration and pricing. As a general partner in a prominent international fund with operations in India put it: "*…Collaboration is fundamental because at the end of the day if you step back and take a look at what a VC's premise is in entering a company, it is generally at very early stages where there is nothing really or very little in the sense of a really stable business - it's fraught with instability… you are really putting your faith in the entrepreneur and in the founding team as such… if you are not sure about collaboration you will always price it in ….*". In a similar vein, a general partner from a prominent local VC, reflecting on collaboration and pricing dynamics asserted: "*So again when I think back, it just reinforces the point that it is all about people and communication… at the end of the day you don't build relationships between spreadsheets, it has to be between people…the entrepreneur is fundamentally irrational…. there have been umpteen cases where deals have broken off at the negotiation stage thinking that you did not give me enough respect for what I bring to the table and you are low-balling me on the price. That kind of thing happens all the time.*". Another general partner from a local VC firm outlined how social attributes matter: "*For collaboration you don't want to be in a situation where half your energy during the collaboration goes towards explaining what you mean. So, a shared vocabulary which* *comes typically from similar interests or similar backgrounds, is important… and of course a shared vision of what they are building is important."*[5]

We propose that boundedly rational VCs foresee potential *benefits* of collaborating with regionally similar or caste-similar founders and reflect the anticipated benefits in how they price the deal. Specifically, in a context rife with uncertainty VCs anticipate that the benefits of better communication and trust will lead to faster, more durable collaboration, and this potential advantage would be reflected in their willingness to pay a higher price (i.e., set a greater pre-money valuation) during the bargaining process. Although founders are also subject to the exact same social forces, as the weaker party in the bargaining process they will benefit more (Tenbrunsel et al. 1999). More formally:

> **H1a:** Greater regional similarity between VCs and company founders is associated with higher pre-money valuation in a financing round.
>
> **H1b:** Greater caste similarity between VCs and company founders is associated with higher pre-money valuation in a financing round.

Although social similarity confers benefits, scholars have also identified the potential costs of homophily. Researchers have noted that socially similar but less diligent actors can impose undue demands, such as tolerance for less effective task performance, exert less than optimal effort (Geertz 1963; Granovetter 1993), or even engage in outright fraud (Yenkey 2017). We posit that VCs anticipate the potential *costs* of collaborating with regionally similar or caste-similar founders and attempt to mitigate them.

An important source of uncertainty for VCs is the potential free riding or lack of optimal effort by regionally similar or caste-similar founders. VCs may rationally react to such potential hurdles to collaboration by increasing their downside risk protection (Bengtsson and Sensoy 2011). In essence, while bestowing regionally similar or caste-similar founders with a higher pre-money valuation, VCs at the same time recognize the costs of

---

[5] Our respondents used words like "background" to denote social similarity markers such as similarity of Indian regions, educational, work experiences, middle-class upbringing etc. However, possibly because it is impolite to discuss caste (akin to race in the US), none of our respondents explicitly noted caste as one aspect of background, although a respondent acknowledged that unconscious bias could nevertheless play a role.

social similarity. Accordingly, they protect their cash flow rights beyond the ownership stake implied by a particular pre-money valuation through greater downside risk protection. More formally:

> **H2a:** Greater regional similarity between VCs and company founders is associated with greater downside risk protection by VCs in a financing round.
>
> **H2b:** Greater caste similarity between VCs and company founders is associated with greater downside risk protection by VCs in a financing round.

The Moderating Effect of Subjective Uncertainty Faced by VCs

Although the venture capital market context is on average highly uncertain, there is still variation in the extent of subjective uncertainty VCs face, depending on the investment stage as well as their prior knowledge of the start-up company's industry. By 'investment stage' we mean the timing of the venture capital investment into companies. Because of uncertainty, VCs typically stage investments in sequential rounds of financing (Amit et al. 1990). The literature commonly identifies seed-/early-stage VC funding as the first professional investment in a start-up company. Subsequent rounds of funding are typically labeled Series A, Series B and so forth. Such sequential investments provide VC firms with a revealed history of the company's operational performance and its prospects over time. In other words, in a later financing round (e.g., a Series B round) a VC has much more objective information on the accomplishments and future potential of a focal company than in a seed-/early-round (Hellmann and Puri 2002). Thus, subjective uncertainty likely decreases over funding rounds.

VCs also vary in the extent of their knowledge of the specific industry in which the portfolio company operates (Dimov and de Holan 2010). VC firms typically acquire expertise in a focal industry from their earlier investments in that domain. When they have a deep knowledge of the particular product market of the focal company, VCs are better able to understand the opportunities and challenges of building a valuable company

(Dimov and Milanov 2010), which reduces their subjective uncertainty. In contrast, when a VC firm has little knowledge of the focal industry, it will experience greater subjective uncertainty.[6]

We propose that while regional or caste similarity is a noisy signal of the ease of future collaboration, the salience of the signal will be greater when VCs face greater subjective uncertainty. Greater salience in turn increases the odds that VCs will weigh regional or caste similarity more when finalizing the pricing decision (Sanders and Boivie 2004). Hence, we expect the main effects of regional similarity or caste similarity on the valuation and downside risk protection components of pricing to be stronger for early-stage deals in comparison to late-stage deals. Likewise, we predict the main effects of regional similarity or caste similarity on pricing to be stronger when the VC firm invests in a company operating in a market niche in which the VC firm has little prior investment history. More formally:

> **H3a:** The effect of regional similarity on pricing (i.e., valuation and downside risk protection) is amplified for early-stage deals compared to late-stage deals.
>
> **H3b**: The effect of caste similarity on pricing (i.e., valuation and downside risk protection) is amplified for early-stage deals compared to late-stage deals.
>
> **H4a:** The effect of regional similarity on pricing (i.e., valuation and downside risk protection) is amplified when the VC firm invests in a market niche where it has little prior investment history.
>
> **H4b:** The effect of caste similarity on pricing (i.e., valuation and downside risk protection) is amplified when the VC firm invests in a market niche where it has little prior investment history.

<u>The Effects of Regional and Caste Similarity on VCs' Financial Performance</u>

Two different theoretical mechanisms suggest a significant relationship between social similarity and financial returns. The first is the establishment (or not) of an effective collaborative relationship (Cable and Shane 1997; Vissa 2011). Irrespective of ex ante expectations, after the investment the VC and founders must collaborate effectively to build a valuable business. Given the high uncertainty and resource scarcity of the situation, building

[6] Because venture capital is a young industry in India, there is much greater variation in the extent of domain specialization among Indian VCs compared to US VCs; this variation makes it an appropriate moderator for our purposes.

a valuable business is inherently conflict prone. Greater regional similarity or caste similarity will facilitate the development of a strong collaborative partnership (Larson 1992) because social similarity objectively establishes the conditions necessary for trust and better communication to emerge and flourish.

The second mechanism is a self-confirming dynamic (Merton 1948) whereby actors enact the very outcome they expect because of confirmation biases. VCs may disproportionately over-allocate their scarce resources (e.g., informal mentoring, expertise and social capital) to companies whose founders have greater regional similarity or caste similarity with them. In doing so, VCs essentially create a self-fulfilling dynamic (Sorenson and Waguespack 2006). Superior collaborative partnerships are in effect created by the VCs' own biased beliefs that lead to biased allocation of the VCs' time and attention. Both these mechanisms suggest:

> **H5a:** VC firms earn a higher financial return on investments in those companies whose founders have greater regional similarity with the VCs.
>
> **H5b:** VC firms earn a higher financial return on investments in those companies whose founders have greater caste similarity with the VCs

METHODOLOGY

Data and Sample

The Indian VC industry was created in 2000 but we restricted our study to the period from April 2005 to December 2012 because in fiscal year 2005-2006 a regulatory change required companies to report detailed information on their shareholding patterns and new equity injections to the government, thus enabling us to accurately compute valuations. Specifically, private firms were required to file periodic electronic returns with the Ministry of Corporate Affairs (MCA) that provided relevant information on shareholding pattern, boards of directors, founders / promoters of the firm as well as balance sheets. More importantly, MCA allowed third parties to download this information for a modest fee[7]. This allowed us and our local data partner Venture Intelligence (analogous to VentureXpert in the US) to construct large databases on relevant private companies. We assembled

[7] MCA charged users a fee of Rs 100 per company(~1.5USD), which allowed users to access the MCA database for a period of 3 hours to download all returns filed by the focal company.

the set of all 1445 VC investments in India during this period and obtained valuation data from government filings for 930 deals, a coverage above 64%. We also added organizational-level data about the VC firms and portfolio companies involved in the deals, and hand-collected fine-grained data about both general partners in the VC firms and company founders. Because of missing data, the final sample for testing our valuation hypotheses comprises 622 funding events of 457 companies (849 founders) by 160 VC firms (471 general partners). For the downside risk protection hypotheses, we had a smaller subsample of 590 deals due to further missing data. Finally, we tested our financial returns hypotheses on the subset of companies that experienced an exit event up to December 2016, which involved 145 VC firm-company observations, of which 44 were successful exits and 101 were failures.

Variables

Dependent Variables

Next we outline how we constructed our three dependent variables: (i) (logged) pre-money valuation of a focal company in a focal financing round, (ii) VCs' downside risk protection in the focal financing round and (iii) (logged) financial return a VC firm obtains from investing into a focal company.

Venture valuation: Consistent with prior research (Bengtsson and Hsu 2015; Hochberg et al. 2010) we define *venture valuation* as the pre-money valuation of a portfolio company at the focal round of funding. We computed valuation in three steps. First, using the regulatory information that companies disclose about the current stock of shares and their price in Indian rupees (INR) as well as number of new shares issued and the price per share, we calculated the deal amount as the number of new shares * price of the new share (which itself is the face value of the share plus possible premiums or discounts). Specifically, we used the PAS3 form that companies filed with the MCA. This form provided raw data on existing and new shareholders, offered and subscribed shares, number and prices of new shares, their face value and discount or premium, specific type of instrument and the conversion ratio for convertibles etc. We then calculated post-money valuation as the [number of outstanding shares + the number of new shares] * new share price. Pre-money valuation is post-money valuation minus the deal amount[8]. The average pre-money valuation in our sample is INR 665 million, (approximately USD 10.9 million[9]) and the median is INR 253 million (approximately USD 4.1 million). We accounted for outliers by winsorizing this variable beyond the extreme 1%.

Downside risk protection: We draw on Bengtsson and Sensoy's (2011) conceptualization of the downside risk protection index to construct our measure of *downside risk protection* from companies' regulatory disclosures[10] as follows. Specifically, we used forms MGT14 and GNL to identify changes in a focal company's Articles of Association (AoA). We created three broad risk categories based on the type of financial instrument used to structure the deal. Some deals were structured as 100% equity through ordinary shares (the riskiest for the VC), some were a mix of both debt and equity through preference shares and convertible debt instruments (medium-risk profile), and some were 100% debt through optionally convertible debt instruments (the least risky for the VC). We subdivided each of these three categories into four subcategories in increasing order of additional contingent cash flow rights that the VC could exercise. The first subcategory consists of investment contracts that stipulate no extra cash flow rights for the VC. The second represents contracts having one of the three following rights: *Cumulative dividends rights*: the VC can compound her dividends until the venture is sold instead of receiving them annually and earn an extra interest on this compounding. The amount due is senior to regular equity. *Redemption rights*: the VC can resell some or all of her shares at predetermined dates for a given price (exercised when the company performs poorly). *Liquidation preferences*: these are a multiple of the original investment (usually above 1) that is paid back upon sale or liquidation, an amount that is also senior to regular equity. The third subcategory comprises contracts having any two of those rights, and the fourth is contracts that have all three.[11] In sum, our measure of downside risk protection is subdivided into 12 ordered categories of increasing VC protection. Our measure of VCs' downside risk protection therefore ranges from a low of 1 to a

---

[8] When the new equity involved more sophisticated deal structures such as funding in multiple tranches, appropriate adjustments were made.
[9] 1 USD = 61.24 INR (as of October 2014).
[10] We don't have access to term sheets. Companies must statutorily report revisions in their Articles of Association (AoA). We can code data only for companies that reflected their term sheet's contractual details through amendments in their AoA.
[11] Bengtsson & Sensoy's (2011) downside risk protection index also includes Participation, Pay-to-play, and Anti-dilution rights. Experts we contacted observed that Participation and Pay-to-play rights were not relevant in the Indian context at the time; Anti-dilution rights were present (by default) on all contracts and hence did not vary in our data.

high of 12 with a mean of 4.7 and a median value of 5. Thus, 50% of the contracts are at most a mix of convertible debt and equity with one additional cash flow right.[12]

Financial returns: We measured the return earned by the VC firm in the investments (cumulated over funding rounds) it made in the focal portfolio company. Prior research that examines the performance of a VC firm's investment into a specific portfolio company uses noisy proxies, such as whether the portfolio company experiences an IPO (Hochberg et al. 2007; Matusik and Fitza 2012). A key strength of our study is that we measure the internal rate of return (IRR) earned by VC firms when investing into a focal company. Theoretically, IRR can range from -100% to infinity (Hawawini and Viallet 2010). An IRR of -100% implies a complete loss of the capital; an IRR of 0% implies the VC firm just recovers its capital; and an IRR of +100% represents a doubling of invested capital in a specific time period – usually a year.

We proceeded as follows. First, we identified the total amount invested by a focal VC firm into a focal company, and then we identified the companies that experienced an exit event, such as an IPO, trade sale or dissolution. Using the exit valuation and the equity structure at exit for these companies (using the same PAS3 form), we computed the payoff received by each VC firm that invested in that company. We then calculated annualized financial return for a focal VC:

$$\textit{Financial return} = [(\text{Payoff} / \text{Total amount invested})^{1/T} - 1] * 100$$

where T is the time expressed in years between the payoff and the investment date. Average IRR (i.e. financial return) in our sample is -28%, with a minimum and median of -100%, and a maximum of 378%. Since this variable was heavily skewed, we shifted it by 101% (to make it positive) so we could compute its logarithm.

Independent Variables

Our key independent variables are *regional similarity* and *caste similarity* between the founders of a focal company and the general partners of the VC firms that invest in the focal company. Consistent with prior research, we defined the founding team as individuals who held significant equity stake in the company (identified based on PAS3 form reporting) and were operationally involved in the company (identified through scrutiny of company

[12] The two components of pricing (pre-money valuation and downside risk protection) are uncorrelated in our data

web sites and media reports). VC firm general partners were identified using the same procedure. Because Indian last names systematically vary with region and caste, we followed prior work (Chen et al. 2015; Vissa 2011) to probabilistically infer individuals' region and caste from their last names as detailed below.

We measure *regional similarity* through language similarity because Indian regions are delineated by specific regional languages. We proceeded in two steps. First, we obtained data[13] from the three largest Indian online matrimonial agencies that maintain details on the last name, caste, religion, and regional language of about six million individuals who are looking for marriage partners. We coded the relative frequencies that a particular last name mapped onto the 14 regional languages of the databases, i.e. *Hindi, Kannada, Malayalam, Tamil, Telugu, Gujarati, Punjabi, Sindhi, Marwari, Marathi, Oriya, Bengali, Mizo,* and *Pahari*. For example, the 27,238 occurrences of the last name "Gupta" in the matrimonial database map to the following languages (relative frequencies in parentheses): Hindi (78.7%), Bengali (7.9%), Punjabi (4.5%), Telugu (2.9%), Marwari (2.5%), and to the nine remaining languages with less than 1%. We interpret these relative frequencies to mean that a person whose last name is "Gupta" is drawn from Hindi-speaking regions with 78.7% probability, from Bengali-speaking regions with 7.9% probability and so forth.

In the second step, we estimated the regional similarity between the founders and the general partners of the VC firm(s) in the focal funding event. For every deal we computed the Mahalanobis distance between the entrepreneurs' and VCs' inferred languages using the 14 languages as axes in a 14-dimensional space. Formally, this distance can be computed as:[14]

$$\text{Regional distance} = \text{sqrt}\left[(\mathbf{LF}_{ENT} - \mathbf{LF}_{VC})^{T} S^{-1} (\mathbf{LF}_{ENT} - \mathbf{LF}_{VC}) \right]$$

Mahalanobis distance corrects for the baseline rate of similarity in the population. The lower the Mahalanobis distance between the inferred languages of entrepreneur A and VC B, the more likely they are from

[13] The information was obtained (after following due process with respect to data confidentiality) as a one-time, raw-data dump. The procedures we employed for data cleanup prior to use in this study are available from the authors.
[14] $\mathbf{LF}_{ENT}$ and $\mathbf{LF}_{VC}$ are vectors representing the frequency positions of each family name of entrepreneurs and VCs, respectively, in a 14-dimensional space of languages. T is the transpose operator and S^{-1} is the inverse of the variance-covariance matrix of the frequency distributions of all last names in our sample across the 14 languages.

the same region. For each deal we considered all the entrepreneurs of the founding team and all the general partners of the VC firms, and used the minimum distance among all these possible entrepreneur-VC dyads.[15] We then multiplied this distance by minus one to generate an increasing measure of *regional similarity* to test our predictions on the drivers of pre-money valuation and downside risk protection. To test our financial returns predictions, we used regional similarity at the moment of first investment of the focal VC firm in the focal company.[16]

C*aste similarity* was computed in the same two-step method as for regional similarity, but in a 5-dimensional space made of the *priest, warrior, merchant, peasant,* and *untouchable* castes. Only 43 of the 1320 unique last names in our sample represented Christian or Muslim religions; we set the caste of these individuals to missing. For each deal we considered all the entrepreneurs of the founding team and all the general partners of the VC firms, and we used the minimum distance among all these possible entrepreneur-VC dyads. We then multiplied this distance by minus one to obtain an increasing measure of *caste similarity* to test our predictions on the drivers of pre-money valuation and downside risk protection. To test our financial returns predictions, we used *caste similarity* at the moment of first investment of the focal VC firm in the focal company. We winsorized *regional similarity* and *caste similarity* beyond the extreme 1% to minimize outlier effects.

Moderating Variables

VC deals can be classified as either early or late stage depending on the state of development of the company receiving funding. Using data from Venture Intelligence we identified start-up or seed-stage deals and expansion or growth-stage deals. Specifically, we classified deals as early stage if the company is raising its first round of institutional financing, is less than five years old and the capital it raised is less than USD 20 million. We classified deals as late stage if the company is raising subsequent rounds of institutional financing and is less than 10 years of age while raising capital. The dummy variable *early stage* is assigned the value 1 if the deal is an

[15] We get the same pattern of findings with alternate variable definitions, such as using only the lead VCs, the general partner on the board of the focal company or restricting the sample to only the first round of funding, significantly reducing sample size. We thank an anonymous reviewer for these suggestions.
[16] We do so because the unit of analysis for the financial returns is not the deal anymore but the VC-company dyad. Hence, measures of similarity cannot change after the first round of funding.

early-stage deal and 0 otherwise. To test H3a and H3b we constructed interaction terms *early stage * regional similarity* and *early stage * caste similarity* after mean centering the similarity variables.

Per deal we also measured the extent to which an industry is new for the VC firm(s). The variable *entering new industry* counts the number of VC firms in the deal that are first-time investors in the industry of the portfolio company. Venture Intelligence classifies companies into industry categories using a classification scheme that maps start-up companies into 24 fine-grained industries. *Entering new industry* can range from 0 to the total number of VC firms in the focal deal. To test H4a and H4b, we constructed the interaction variables *entering new industry * regional similarity* and *entering new industry * caste similarity* after mean centering the variables[17].

Control Variables

We first specify the controls in our analyses of the drivers of *venture valuation* and *downside risk protection*. At the level of the venture, we counted the *number of prior funding events* which counts the number of times a venture received institutional money prior to the focal deal. As a correlate of quality, we expect this indicator should increase valuation and so should *venture age*. We proxied the founding teams' quality by the eliteness of the educational institutions the founders attended. *Founding team eliteness* counts the number of undergraduate and post-graduate degrees obtained at prestigious institutions by all the members of the founding team. Total *start-up experience* is a count of the number of prior ventures founded by all the members of the founding team. Since this signals prior related experience, it should drive up valuation. Finally, since VCs may unconsciously associate the entrepreneurs' merchant caste membership with greater business acumen, we control for the *proportion of merchant caste* in the founding team.

We also control for VC firm attributes: the *number of board seats* taken by the VCs in each deal, *VC experience* which counts (scaled in tens) the number of prior deals the VC firm(s) already executed before the focal deal, and *VC firm status* which is the sum of the normalized (Bonacich 1987) centralities in the syndication network of the VC firms in the deal. The syndication network was constructed using the VC firms' syndication alliances over five-year moving windows prior to each deal date (Ozmel et al. 2013; Sorenson and Stuart 2001)

[17] An alternative measure constructed at the individual General Partner level of analysis yielded the same pattern of findings. We thank an anonymous reviewer for this suggestion.

in our dataset, which includes all VC firms (purely local, purely foreign as well as foreign firms with local funds). Higher centrality in the syndicate network indicates higher VC firm status in our specific setting.

On a macro level, we control for *VC competitive intensity* with the number of VC firms located within a 150-kilometer radius of the focal company at the time of funding since Sorenson and Stuart (2001) find this distance is a critical cutoff. We calculated *industry hotness* as the average deal size within the industry of the start-up company in the six months prior to the deal date. To account for spatial proximity effects, we computed the *geographical distance* (in thousands of km) as the minimum of the distance between the headquarters of the company and any Indian headquarters of all the funds in each deal. In addition, we included *industry dummies* to control for the 24 fine-grained industries that Venture Intelligence uses to classify companies, *region dummies* to account for the region of the company, *investor type dummies* to control for whether the deal involves only purely foreign VCs with India-dedicated funds, local funds, or a mix of types, and *year dummies* to account for year-specific shocks.

Finally, past research suggests that several parameters related to the VC firm, the venture, and the industry should affect the VCs' *financial returns*. Concerning the VC firm, we controlled for VCs' attention to the focal venture by counting the *number of ventures* (scaled as tens of ventures) the VC firm was invested in at the moment of its first investment into the focal company (Buchner et al. 2017; Ljungqvist and Richardson 2003). We approximated VCs' commitment by the total *number of rounds* the VC firm invested in the company until exit (Buchner et al. 2017), and the *total amount* the focal VC firm invested (across rounds) in the focal company (Buchner et al. 2017; Ljungqvist and Richardson 2003). We also accounted for *VC firm status* (Nahata 2008). Regarding the venture, Sahlman (1990) notices a link between VCs' returns and the venture's quality and performance. Since sales information is not available, we proxied venture quality by *founding team eliteness, start-up experience* and *proportion of merchant caste* in the founding team. Concerning the industry, we included an indicator variable for the information technology industry (*IT*) since about half the sample comprises companies from this industry (Amit et al. 1998), a company *region dummies* as well as the *exit year* of the company (Cochrane 2005). In addition, to proxy monitoring and assistance we included the *geographical distance* between the VC firm and the company (Manigart et al. 2002). Finally, we introduced our other two dependent variables: *first* *investment valuation,* the pre-money valuation of the company at the time of first investment by the VC firm, as well as the *downside risk protection* at the time of first investment.

Analysis Approach

The VC market is a two-sided matching market in which the performance outcomes of VC firms and their portfolio companies depend on both sorting (VCs and entrepreneurs mutually decide to enter into an investment relationship) as well as collaboration (VCs actively coordinate and cooperate with entrepreneurs subsequent to investment), in order to build a valuable venture (Sørensen 2007). Because sorting may occur on unobserved characteristics, such as entrepreneurs' quality or VCs' willingness to let entrepreneurs learn from execution errors, measures of social similarity (our independent variables) could become endogenous in the regression analyses of pricing or returns if, for instance, sorting occurs such that higher quality entrepreneurs match with socially similar VCs. We accounted for this endogeneity by triangulating across two different methods. Obtaining consistent findings across these procedures increases our confidence that observed patterns in our data are indeed driven by our hypothesized relationships. Our first approach is a classical, two-stage regression with a Heckman selection model (1979). Our second method utilizes Sørensen's (2007) two-sided matching Bayesian estimation model implemented in R using libraries developed by Klein (2016).

Two-stage Heckman Selection Approach

We assume VC firms first select a company from a promising set and then determine their pricing (both valuation and downside risk protection) for that company. In the first stage we model the probability of the selection of the focal company for funding by the focal VC firm using a choice-based, sampling approach (Sorenson and Stuart 2001) to construct a counterfactual set of deals. Specifically, of all the funds that invested in a period of six months prior to the deal, we paired each one at random with a company they had *not* funded. We constructed this pool of counterfactual VC firms and companies per number of prior funding events to mitigate quality differences across companies as a driver of selection. Thus, if a company had received two rounds of prior funding then its counterfactual set was built from only those VC firms that had funded companies who themselves had already received two rounds of prior funding. We chose a six-month interval because it is the usual time to obtain venture financing in India. Using this methodology, 4849 counterfactual deals were created which gives a ratio of about 3.5 to 1 of counterfactual to real deals.

For the selection equation we used the instruments *median regional similarity* and *median caste similarity* at the state–industry–year level for each venture, where state refers to the Indian state where the venture is located, industry is the industry the venture belongs to (as described above) and year is the year of the focal deal. These instruments satisfy the exclusion-restriction principle since median similarities should affect the overall likelihood of funding based on social similarity, but should not affect the pricing conditions of any single venture per se (Hegde and Tumlinson 2014). For these counterfactual deals, we recreated the variables *number of prior funding events*, *founding team eliteness*, *start-up experience*, *proportion of merchant caste, geographical distance*, *industry hotness*, *regional similarity*, *caste similarity,* as well as the dummies described above. We also constructed a *new tie* dummy variable that measures whether the VC firm was investing in the company for the first time since this should affect the likelihood of obtaining funding (Hallen 2008). We introduced the inverse Mills ratio as a control (Heckman 1979) in the second-stage valuation equation as well as in the second-stage downside risk protection equation.

Bayesian Two-Sided Matching Approach

Sørensen's (2007) methodology enables us to explicitly focus on a two-sided matching process in which the entrepreneur and VC mutually choose to work with each other (or not). VC firms and entrepreneurs mutually select who to pair up with based on observable criteria, such as status of VC firms, prior experience of entrepreneurs etc., as well as unobservable factors related to other competing actors in the market. In the first stage, the key identifying assumption is that the presence and features of the other competing actors (namely, other VC firms competing with the focal VC to provide funding or other start-up companies competing with the focal start-up company to receive funding) are exogenously given. We use this approach to test our hypotheses on the drivers of financial returns (H5a and H5b). Like the Heckman selection model, the first stage in our method is a discrete choice model – but one that allows for the interaction of choices made by *different* actors. The data-generating process assumed in this modeling approach is the Gale et al. (1962) two-sided matching model known as the college admissions model. The second stage is the outcome equation, which specifies the focal VCs financial returns.

Because a given actor's investment decision interacts with that of all other actors, conventional estimation is computationally infeasible. Instead, we utilize Sørensen's (2007) Bayesian estimation methods (Markov Chain Monte Carlo (MCMC) simulations), with a prior distribution that is as uninformative as possible while maintaining a proper posterior distribution – which is informed by the observed data. Specifically, in the first-stage model the dependent variable is the pairing of a focal VC firm investing in the first round in a focal portfolio company. The independent variables in the first and second stage are the exact same variables as in the Heckman selection model. The dependent variable in the second stage of the Bayesian estimation model is the financial return experienced by the focal VC from the investment in the focal company. Triangulating these results with an OLS of financial returns on our similarity variables yielded similar findings.

RESULTS

Tables 1-3 report all the descriptive statistics and correlations for our analyses. Bivariate correlations between all the right-hand variables are moderate to low; further, low VIFs (maximum of 1.6) suggest that multi-collinearity is not a concern. A typical deal is the venture's first (only 25% of the ventures receive more than one round of funding), it is an early-stage deal (66% are early stage) and is funded by a single VC firm (74% of the deals are not syndicated) located less than 29km from the venture (the median of geographical distance is 29km). The median deal amount is USD 1.7 million at a pre-money valuation of USD 4.1 million.

Insert Tables 1a, b & 2 & 3 about here

Heckman Selection Approach: First-Stage Estimates

Table 4 displays the results of the probit modeling of the likelihood of a company receiving VC funding. In all tables and for all variables we report the results of one-tailed t-tests. As expected, *founding team eliteness* increases the likelihood of receiving funding (p-value<0.001). Consistent with Sorenson and Stuart (2001), greater *geographical distance* decreases funding likelihood (p-value<0.001). The negative coefficient of *new tie* (p-value<0.001) means that not having been funded before by the VC also decreases the likelihood of ever being funded by that VC (Hallen 2008). Also, consistent with prior work (Vissa 2011), *regional similarity* is positively related to receiving funding (p-value=0.046). Both median regional similarity and median caste similarity at the state-industry-year level are negatively associated to funding probability (p-value=0.003 and 0.06) because funding mainly occurs in big cities where individuals are more dissimilar on average.

Insert Table 4 about here

Heckman Selection Approach: Second-Stage Valuation Estimates

In Table 6 we report the results of the second-stage valuation regressions. We clustered standard errors both by VC firm and company using two-way clustering (Cameron et al. 2011; Petersen 2009) to account for non-independence of multiple observations drawn from a focal VC firm or company. Model 1 represents the baseline model. In Model 2 we introduced our independent variables *regional similarity* and *caste similarity* to test H1a and H1b. *Regional similarity* is significant (p-value=0.001) and positive, which suggests that greater regional similarity positively impacts valuation and thus strongly supports H1a. In contrast, *caste similarity* is negative and significant (p-value = 0.022) contradicting our theoretically derived prediction H1b.

Insert Table 6 about here

In Model 3 we test whether the effect of similarity on valuation will be accentuated for early-stage deals as compared to late-stage deals (H3a and H3b). *Early stage * regional similarity* is positive and significant although at the 10% threshold (p-value=0.064 and 0.071 in Model 3 and the full Model 5, respectively), which lends some support to H3a. *Early stage * caste similarity* is positive but non-significant at conventional levels, which does not support H3b. Next, in Model 4 we test whether the effect of similarity on valuation is amplified when VCs invest in industries that are new to them (H4a and H4b). The coefficient of *entering new industry * regional similarity* is positive and significant (p-value=0.002 in Model 4 and 0.003 in the full Model 5) thereby supporting H4a. However, the coefficient of *entering new industry * caste similarity* is negative and significant (p-value<0.001 in Models 4 and 5, respectively), which contradicts our theoretically derived prediction H4b.

Our main effects are also practically relevant. An increase of one standard deviation in regional similarity yields a rise in valuation of USD 195,000 which represents a 1.8 % increase above the mean pre-money valuation and a 4.7% increase above the median pre-money valuation. Likewise, an increase of one standard deviation in caste similarity decreases valuation by USD 180,000 which represents a 1.6% drop from the mean pre-money valuation and a 4.3% decrease from the median pre-money valuation.

Other control variables that are robust across specifications also merit our attention. In line with expectations, as ventures progress in their development they tend to receive higher VCs' valuations which is illustrated by the strongly significant (p-value at most 0.001) and positive coefficient of *number of prior funding events* and strongly significant (p-value at most 0.001) and negative coefficient of *early stage*. In addition, the level of prior *start-up experience* of the entrepreneurial team is also positively valued by VCs (p-value at most 0.005). Surprisingly, however, we find that contrary to prior work studying the US (Hsu 2004), *VC firm status* is positively and significantly (p-value at most 0.001) associated with higher prices paid for venture equity.

Heckman Selection Approach: Second-Stage Downside Risk Protection Estimates

Table 7 presents the results of the second-stage downside risk protection regression, using the same variables we used for the valuation analysis. Since our dependent variable is an almost continuous index varying from 1 to12, we employed an OLS specification for ease of interpretation. Our results are, however, robust to ordered logit and negative binomial specifications as well.

Insert Table 7 about here

The coefficient of *regional similarity* is positive and significant across all models (p-value<0.02 in Models 2-5) which lends strong support to our prediction (H2a) that VCs attempt to mitigate potential free-riding by regionally similar founders through better downside risk protection. However, while the coefficient of *caste similarity* is positive, it is not significant at conventional levels, suggesting no support for H2b. Models 3 and 4 test the same interactions as in the valuation regression and Model 5 reports the full model. Unfortunately, none of the interactions (early stage or entering new industry) with regional similarity or caste similarity are statistically significant.

Overall, our analyses of the drivers of both the valuation and downside risk protection components of pricing suggest that VCs set greater pre-money valuation for companies led by regionally similar founders, and at the same time mitigate their risks through greater downside risk protection for these companies. However, caste similarity seems to have a theoretically anomalous effect because greater caste similarity drives down valuations.

Financial Returns to Venture Capital Firms

Panel A of Table 5 reports OLS results on the drivers of IRR. Model 1 shows estimation results with only the control variables. Given the paucity of prior studies examining VCs' financial returns on investments into specific companies, we first focus on these results. Higher status VCs earn more on their investments (p-value of VC firm status=0.024); however, when we account for our independent variables in Model 2 this effect disappears (p-value =0.125). We also observe that investments into companies in the IT industry are less profitable (p-value of *IT* = 0.046). Interestingly, the *number of rounds* a VC fund has invested in a venture is positively related (p-value<0.001) to the financial performance of the VC, which indicates that VCs do a good job of follow-on investments in the most promising ventures. Further, *first investment valuation* is positive and significant (p-value at most 0.01 through Models 1 and 2 and at most 0.028 across all models) which suggests that initial valuation matters for eventual IRR, but in a nuanced way. VCs seem to build more valuable ventures in partnership with founders rather than (merely) select the most promising start-ups.

Insert Table 5 about here

In Model 2 we test whether similarity influences VCs' financial returns. The coefficient of *regional similarity* is positive and significant (p-value=0.025) indicating that greater regional similarity yields better financial performance and providing strong support for H5a. The impact of regional similarity on financial returns is also practically relevant. Indeed, holding all variables at their mean and increasing *regional similarity* by one standard deviation improves average IRR from -28.22% to -26.59%. This rise of 163 basis points represents 5.8% of the average IRR. However, the coefficient of *caste similarity* is negative and just beyond conventional significance thresholds (p-value=0.107), contradicting our theoretically derived prediction H5b.

Bayesian Two-sided Matching Model of Financial Returns to Venture Capital Firms

Model 3 of Panel B in Table 5 reports our results using Sørensen's (2007) Bayesian two-sided matching approach. This analysis uses a smaller sample (n=138) compared to Panel A (n=145) because we drop seven observations in order to exactly replicate Sorenson's (2007) approach. To facilitate a direct comparison between Bayesian and OLS approaches, we juxtapose our Bayesian analysis in Model 3 with OLS results in Model 4 where the OLS is run on the exact same sample (n=138) used for the Bayesian analysis reported in Model 3. The lower part of Model 3 displays the Bayesian estimates of the matching equation (akin to the selection part of the Heckman model), and the upper part of Model 3 reports the Bayesian estimate of the financial returns equation (akin to the second-stage Heckman model) correcting for the effect of unobserved variables on sorting, such as unobserved entrepreneurs' quality or VCs' willingness to let entrepreneurs learn from execution errors.

First, we note that the Bayesian and OLS coefficient estimates of the drivers of IRR (Models 3 and 4, respectively) are broadly consistent. In Model 3 we see that total *number of rounds* a VC has invested in a venture significantly improves the VC's returns (p-value<0.001) and so does the *first investment valuation* (p-value=0.028 in Model 3). The coefficient of *regional similarity* is positive and significant (p-value=0.025 in Model 3) further lending support to H5a. However, *caste similarity* is negative and significant (p-value=0.056 in Model 3), which contradicts our theoretically derived prediction H5b. Combining these results with the OLS results of Panel A, we conclude that our prediction on regional similarity (H5a) is supported by the data. However, there is evidence that the effect of caste similarity (H5b) is theoretically anomalous with greater caste similarity decreasing financial returns.

Robustness Tests and Additional Analyses

Our results are robust to several alternative specifications and measures. First, we examined whether our measurement strategy of inferring caste and regional similarity from last names was driving the results. Specifically, our measures of regional similarity and caste similarity stem from the "fuzzy" (varying from 0 to 1) assignment of a person to a particular caste or region, rather than a crisp 0 or 1 assignment. To test robustness of our results to this measurement procedure, we transformed our fuzzy measure of social similarity into a crisp measure (based on the main mode of the distribution) and obtained the same pattern of findings.

Another concern is the possibility that higher quality founding teams would receive multiple offers from different VCs and could (unconsciously) choose the VC that is most similar in terms of region. Therefore, regional similarity would be a correlate (but not a cause) of valuation, since unobserved venture quality is the causal driver. There are two reasons why we believe this is less likely to be operating in our data. First, our qualitative interviews suggest that our study period represents a cold VC market when an entrepreneur receiving multiple terms sheets from competing VCs was a very rare event. Second, if this mechanism were indeed operating in our data, we should expect a positive and significant interaction effect between observable indicators of venture quality and regional similarity on valuation. However, we find that the interaction terms *founding team eliteness * similarity* and *start-up experience * similarity* are not significant at conventional levels.[18]

Unpacking the Theoretically Anomalous Effect of Caste on Pricing and Returns

Our surprising finding that greater caste similarity decreases valuation and returns is theoretically anomalous and contradicts our conceptual framework that focuses on the benefits of social similarity. On the contrary, our results suggest that caste *dissimilarity* is associated with higher valuations and returns. One source of the anomaly could stem from our conceptualization of caste as a nominal social category. To investigate the anomaly of caste-dissimilar matching we relaxed the assumption of caste as a nominal category; we instead treated

[18] Allaying unobservable heterogeneity concerns through Bayesian analysis on the drivers of valuation (or downside risk protection) is not feasible because Sorenson's (2007) two-sided matching model assumes that both initial selection and pricing (both valuation and downside risk protection) occur in a single (sorting) step, and the model focuses on how sorting and subsequent treatment (i.e., collaboration) drives performance outcomes (i.e., IPO).

caste as a rank-ordered social category and probed the implications. Specifically, we considered caste membership as a rank ordering (from highest to lowest) of priest, warrior, merchant, peasant, and untouchable categories.

Recalling that VCs wield economic power over founders, we can discern two types of matching between VCs and founders that could explain our observed findings on caste. Higher-caste VCs can match with lower-caste founders which furnishes the VC with both economic and social power over the founder. This type of caste-dissimilar matching is hierarchically consistent since the social power difference reinforces the economic power difference between the two actors. Alternatively, lower-caste VCs can match with higher-caste founders. This type of caste-dissimilar matching is hierarchically inconsistent because the social power differences conflict with the economic power difference between them; the VC has economic but not social power.

There are two competing explanations that underlie hierarchically consistent matching whereby higher-caste VCs match with lower-caste founders. On the one hand, VCs may pay more for caste-dissimilar founders only when such founders signal exceptional quality, in other words for purely rational reasons. On the other hand, VCs may pay more for caste-dissimilar founders due to nonrational feelings of "*noblesse oblige*" (Homans 1950), that is, high-status actors feel obligated to help low-status actors. Both explanations imply that higher-caste VCs will set higher valuations for lower-caste founders although the former explanation also implies this tendency to set higher valuations will be particularly pronounced when lower-caste founders signal superior quality through credentials such as graduating from elite educational institutions like the IITs or IIMs.

In addition, if access to deal flow in Indian VC markets is stratified by caste, we expect to find patterns of hierarchically inconsistent matching of lower-caste VCs with higher-caste founders in our data. Specifically, if access to deal flow in Indian VC markets is stratified by caste, then we expect lower-caste VCs to set higher valuations that benefit higher-caste founders because doing so improves future deal flow quality for low social rank VCs, a mechanism noted by Zhang et al. (2016) among ethnic-Asian VCs willingness to set higher valuations for white-ethnic-led companies in Silicon Valley.

<u>Measuring caste dissimilarity</u>: To measure caste dissimilarity where caste is treated as a rank-ordered attribute, we first performed the same two steps as in our analysis of caste similarity in the 5-dimensional space composed of the priest, warrior, merchant, peasant, and untouchable castes, but we did not multiply the distance

by minus one to insure an increasing measure of dissimilarity. Then, in a third step to account for the rank ordering of caste, we splined this variable into three parts: *Caste dissimilarity (VC<ENT)*, *Caste dissimilarity (VC>ENT)*, and *Caste dissimilarity (VC=ENT)*. *Caste dissimilarity (VC<ENT)* was assigned the value of caste dissimilarity when the caste superiority index (described in the next paragraph) was >0.05 and 0 otherwise; *Caste dissimilarity (VC>ENT)* took the value of caste dissimilarity when the caste superiority index was <-0.05 and 0 otherwise; finally, *Caste dissimilarity (VC=ENT)* was defined as the value of caste dissimilarity when the caste superiority index was between -0.05 and 0.05 and 0 otherwise.[19] Using these variables we tested our predictions on the drivers of pre-money valuation as well as on the downside risk protection of VC firms. To test our predictions on the drivers of financial return, we considered these same three caste dissimilarity variables but at the moment of first investment of the VC firm into the focal company. We winsorized caste dissimilarity beyond the extreme 1% to minimize the effects of outliers.

<u>Caste superiority index:</u> Figure 1 details the computation of this index through specific numerical examples. For each entrepreneur and VC, we wanted to assign a single numeric score that indicates the "value" of their caste distribution. Since technology entrepreneur or VC is a socially desirable position in the Indian economy, there should be an overrepresentation of higher castes with respect to untouchables and that ratio of overrepresentation in this context suggests how much better it is to be a priest or a warrior than an untouchable. To compute those weights, we generated the natural distribution of the five castes among the entrepreneurs and VCs in our sample by summing across the caste frequencies and normalizing by the total number of people. We then computed a ratio of the occurrence of each of the first four castes to untouchables. For instance, in our sample, for each untouchable there are on average 3.9 priests, 5.3 warriors, 3.1 merchants, and 3.1 peasants. To calculate the caste score we added the products of the caste distribution frequencies of each person with those weights. Then for each entrepreneur-VC dyad in a given deal, we subtract the caste value of the VC from that of the entrepreneur.

[19] Since our measure of (dis)similarity is a continuous distance variable based on a probabilistic caste distribution, there are no cases with a pure 0 distance (exact similarity) unless the entrepreneurs and VCs have the exact same names. Therefore, we make a cut-off at |0.05| for the caste superiority index since the case where it is 0 is very rare. As it is, there are roughly 8% of deals in that category, and all our results are robust to different cut-off values ranging from 0 to |0.05|.

Caste superiority index is the average of these differences over all the possible dyads of the deal. When this variable is positive the entrepreneurial team tends to be of higher caste than the VC team and vice versa.[20]

Insert Figure 1 and Table 8 about here

Using the same methodology as in Table 6, we report in Table 8[21] the results of the second-stage valuation regressions with our splined *caste dissimilarity* variables. Model 1, the baseline, is identical to Model 1 in Table 6. In Model 2 we introduce our independent variables *regional similarity* and the three *caste dissimilarity* variables to test their effect. *Regional similarity* is significant (p-value=0.001) and positive, which provides evidence that greater regional similarity positively impacts valuation and thus still strongly supports H1a. In line with our reasoning, *caste dissimilarity (VC>ENT)* is positive and significant (p-value=0.008), indicating that caste dissimilarity is associated with greater valuation when the caste of the VC team is higher on average than the entrepreneurial team's – although the precise mechanism (*noblesse oblige* or intrinsic quality) is still unclear. *Caste dissimilarity (VC<ENT)* is not significant, suggesting that unlike the ethnic-Asian VCs in Silicon Valley observed by Zhang et al. (2016), lower-caste Indian VCs do not accrue functional benefits in matching with higher-caste founders by offering a higher valuation.

In Models 3 and 4 of Table 8 we test whether the greater valuation that higher-caste VCs set for companies led by lower-caste founders is accentuated for early-stage deals as well as when VCs invest in industries that are new to them; Model 5 reports the full model. In Model 3, the coefficient *early stage * caste dissimilarity (VC>ENT)* is not significant, suggesting the anomalous effect of caste is not accentuated for early-stage deals. In Model 4, the coefficient *entering new industry * caste dissimilarity (VC>ENT)* is positive and significant (p-value= 0.018), evidence that the anomalous effect of caste is accentuated when VCs enter into new industries. Overall, we conclude that the theoretically anomalous effect of a negative impact of caste similarity on valuation is driven mainly by hierarchically consistent caste- dissimilar matching when the caste of the VCs is higher than the

[20] An alternate measure of the caste superiority index where priest, warrior, merchant, peasant and untouchable are coded respectively as 5,4,3,2 and 1 yielded the same pattern of results.

[21] To conserve space, with the new *caste dissimilarity* variables, we only report the results of valuation in Table 8. The results of the selection equation, the downward protection index and the drivers of IRR remain unchanged and are available upon request from the authors.

entrepreneurs'. In other words, when VCs are higher caste relative to the entrepreneurs in a focal deal, the VC sets a higher pre-money valuation, and this effect becomes even stronger when VCs invest in a new industry.

To adjudicate between the two competing explanations, *noblesse oblige* and intrinsic quality, that underlie hierarchically consistent matching, we explored whether founding-team eliteness moderated the main effect of caste dissimilarity. As we can see in Model 6 of Table 8, *founding team eliteness * caste dissimilarity (VC>ENT)* is positive and significant (p=0.085), suggesting that higher-caste VCs' propensity to set higher pre-money valuations for lower-caste entrepreneurs is accentuated when those founders are graduates of elite educational institutions. These results are consistent with the intrinsic quality mechanism rather than *noblesse oblige*.

Further, in unreported results, we found that hierarchically consistent matching between VCs and entrepreneurs has the same effect on the drivers of IRR. Overall, we conclude that conceptualizing caste as a rank-ordered social attribute and focusing on more complex social processes around dissimilarity may be a promising approach to reconcile our surprising empirical results on caste similarity - findings that contradict well-understood theoretical mechanisms that highlight the benefits of social similarity.

DISCUSSION AND CONCLUSION

Our results on the effects of regional similarity, conceived as a nominal social attribute, are theoretically straightforward. They paint a clear-cut picture of boundedly rational VCs projecting into the future to anticipate the benefits and costs of regional similarity in building a collaborative partnership and reflecting these benefits and costs in the deal pricing through respectively setting higher pre-money valuation but at the same time mitigating their downside risks using contractual terms that reflect a more debt-like claim on future cash flows. Further, we find that greater regional similarity is associated with higher financial returns for VCs – which is consistent with both a 'correct anticipation' explanation as well as an alternative mechanism of biased allocation of VCs resources that fuels a self-fulfilling dynamic.

However, our results on caste similarity are theoretically anomalous. Contrary to our conceptual framework, we found that VCs set higher pre-money valuations for companies led by caste-dissimilar founders. We delved deeper in post-hoc analyses using a rank-ordered conceptualization of caste. We found that VCs set higher pre-money valuations for companies led by founders who have lower caste rank than themselves. Interestingly, this effect was stronger when the lower-caste founders graduated from elite educational institutions, an important marker of founders' intrinsic quality. Overall, these findings paint a different and more ambiguous picture of how caste influences VC behavior and outcomes, one of boundedly-rational VCs being influenced by caste dissimilarity rather than similarity. We note that our findings are based on observational data, using a novel caste and language coding methodology that may suffer from unknown bias and we do not observe the processual mechanisms (Clough et al. 2019) underlying our findings. Despite these limitations, as we discuss below, our overall pattern of results advances the venture capital literature in several ways, in addition to offering potential research pathways for organizational science in general.

How do social forces matter for VC behavior?

We harmonize theoretical ambiguities around how social forces matter for VC behavior, clarifying a set of somewhat inconsistent prior findings in the literature. Prior research examines the sole social attribute of ethnicity and conceptualizes ethnicity as either nominal (Bengtsson and Hsu 2015; Hegde and Tumlinson 2014) or rank-ordered (Zhang et al. 2016), reflecting the broader lack of scholarly unanimity on whether ethnicity is nominal or rank-ordered in the US context. Considering the question 'How does social similarity influence VC behavior?' our study suggests that the first step is to examine the "rank-orderedness" of the focal social attribute. Scholars in the status characteristics paradigm (Berger et al. 1980; Ridgeway 1991) show that ascribed nominal social category attributes such as gender, age, race and ethnicity often become rank-ordered status attributes because individuals in a specific setting learn to associate different personal characteristics with general ability due to socialization in the broader culture. For example, researchers have shown that it is a widely held assumption in contemporary American society that it is more worthy, worthwhile or superior to be male rather than female, or European American rather than other ethnicities (Berger et al. 1980; Eagly and Wood 1982).

Whilst purely nominal attributes would likely trigger the well-studied similarity (homophily) mechanism, the more a social attribute becomes infused with hierarchical considerations and hence rank-ordered, the more likely it is that more intricate dissimilarity-based mechanisms may influence behavior. We suggest that rank-orderedness of a social attribute matters because it brings power into the equation when discussing similarity, which is absent when we conceptualize social attributes as purely nominal categories.

Our finding that regional similarity (a purely nominal social attribute in our empirical context) has a positive effect on valuation is concordant with Bengtsson and Hsu's (2015) empirical results that US VCs pay more for co-ethnicity and consistent with Hegde and Tumlinson's (2014) theoretical model that emphasizes the benefits of social similarity. At the same time, contrary to Bengtsson and Hsu's empirical results that point to non-rational bias, we find that Indian VCs set more onerous contractual terms for regionally similar founders, suggesting that unlike the US VCs, Indian VCs are mindful of the costs of social similarity and are rationally attempting to mitigate those downside risks. Therefore, our findings suggest that examining only company selection or valuation, without also examining drivers of VC contract design (Kaplan and Strömberg 2003; Sahlman 1990) may not present an accurate picture because VCs' downside protection is an important risk mitigation mechanism that seems to be regulated by social forces.

Likewise, our result that VCs pay more when founders are caste-dissimilar (conceptualizing caste as a rank-ordered social attribute) is consistent with Zhang et al. (2016)'s empirical focus on dissimilarity in social attributes. We place Zhang et al.'s (2016) interesting empirical results on firmer theoretical footing by providing a more complete conceptual model of how dissimilarity on rank-ordered social attributes drives VC behavior. Zhang et al. examine one specific type of dissimilarity; low-status VCs matching with high-status founders. They propose that low-status VCs may rationally pay more for high-status founder led companies because it improves low-status VCs' future deal flow. We broaden Zhang et al.'s (2016) focus and consider both types of dissimilar matching; low-status VCs matching with high-status founders as well as the obverse. We provide and empirically tease apart competing theoretical explanations for the obverse situation wherein high-status VCs pay more for low-status founder led companies. Specifically, we adjudicate between a "*noblesse oblige*" explanation that assumes an altruistic orientation by high-status VCs and an "intrinsic quality" explanation that implies a (boundedly) rational orientation by VCs. Our finding that VCs pay particularly more for caste-dissimilar founders with elite education implies that the "intrinsic quality" mechanism is operating in our data, suggesting VC behavior is more consistent with assumptions of bounded rationality than altruism.

Indeed, our above findings on dissimilarity combined with our results that VCs pay more for regionally similar founders while at the same time mitigating their downside risks suggests that VCs' behavior may be more (boundedly) rational than previously believed and opens opportunity to extend the literature on VCs' decision biases. This literature (Franke et al. 2006; Zacharakis and Shepherd 2001) focuses primarily on the selection stage to study how unconscious social biases influence VCs' initial evaluative processes. Our findings imply that when pricing is seen more holistically to include both valuation and downside risk protection, these biases are somehow mitigated at the deal pricing stage. We advocate for richer ethnographic and inductive studies on company selection and deal negotiation to bridge this gap in knowledge. Inductive theorizing on the role of social forces in the mutual evaluations of VCs and founders and its impact on company selection and deal pricing can shed useful light on venture capital firms' organizational routines and practices through which general partners' rational and non-rational motivations may regulate pricing outcomes in VC markets.

Zooming out, the broader management literature on homophily in work teams (McPherson et al. 2001), set mainly in the Western world, usually assumes that the demographic attributes on which actors are similar are nominal. Very little work acknowledges that some demographic attributes might be rank ordered. A rare exception is research by Mitsuhashi and Greve (2004) that explicitly models organizational tenure differences among members of top management teams as a rank-ordered social attribute infused with power. Complementing their work on intra-team power differences, we examine inter-team power differentials and model how dissimilarity in caste rank between VCs and founders influences behavioral and performance outcomes.

Interestingly, the broader management literature that does focus on status ranking overwhelmingly focuses on status homophily (Sauder et al. 2012). Relatively little attention is paid to status heterophily – specifically how and why high-status actors exchange with low-status ones. Among the explanations given are that high status actors may prefer to exchange with low status ones because the latter are more likely to exert greater effort (Castellucci and Ertug 2010) or because the former's poor performance against aspirations makes them more prone to making risky choices in partner selection (Shipilov et al. 2011). We add to these explanations by suggesting that high-status actors may agree on terms of exchange that are more favorable to low-status actors when the latter signal greater intrinsic quality. More generally, our empirical context provides an opportunity to study rank-ordered social attributes such as caste that could play an increasingly important role in economic action as India becomes a more central constituent of the global economy (Bapuji et al. 2019).

The role of caste as a social force influencing economic exchange in India

Caste is an important social force in the context of economic action within and between large, publicly listed companies and business groups in India. For example, scholars have shown that caste similarity between equity analysts and the CEO of their target firms regulates equity analysts' forecast accuracy (Chen et al. 2015); CEO compensation is influenced by caste similarity between the Chair and CEO (Damaraju and Makhija 2018); caste similarity is also a significant driver of the composition of boards of directors of public firms (Bhagavatula et al. 2019) and mergers and acquisitions among large, established firms (Bhalla et al. 2019). Likewise, caste is important for economic exchange in traditional entrepreneurship (Menning 1997).

In contrast, the role of caste in the burgeoning, high-growth, technology intensive entrepreneurship sector of the Indian economy is unclear. Vissa (2011) provides evidence that entrepreneurs in this domain use caste similarity as a filter in their *intentions* to form new interpersonal network connections with potential customers or alliance partners; but caste similarity was not a significant predictor of the *realized outcome* of actual inter-firm economic exchange tie formation. Like Vissa (2011), our study samples entrepreneurs and VCs from India's burgeoning middle class that is located mainly in the big cities, is highly educated, and operates or funds growth ventures. Consistent with Vissa (2011) , we find caste similarity is not a significant driver of company selection in our sample of realized VC deals.

In addition, when caste is measured as a rank-ordered variable (cf. Table 8), we find that caste similarity does not predict pricing outcomes. Rather we find evidence that higher-caste VCs set higher valuations for lower-caste founder led companies, particularly when such founders signal intrinsic quality through elite education credentials. Further, unlike Zhang et al.'s (2016) findings that imply deal flow stratification by ethnicity in Silicon Valley, the lack of a significant effect for lower-caste VCs matching with higher-caste founders implies that deal flow is likely not stratified by caste in our sample. We conjecture that there could be two broad reasons for this. First, perhaps caste categories in urban India have become broader such that individuals from the priest, warrior and merchant castes, the pre-dominant ones in our sample, do not view themselves as belonging to separate caste categories anymore. Second, other foci such as education or work organizations, rather than caste, could have become the new anchor around which network ties coalesce (Chen et al. 2015) for the actors in our sample who are predominantly young(er). Clearly more research is needed to understand how VC deal flow is stratified in the Indian entrepreneurial ecosystem. Perhaps an important mechanism is syndicate formation. Examining the caste dissimilarity drivers of syndicate formation in Indian VC deals and comparing and contrasting it with syndicate formation in Indian investment banking deals to determine how the effects of caste varies in these two domains is an important area of future research.

Our puzzling non-results of caste similarity's effect on pricing doubtless warrant more research. One explanation could be that the costs of social similarity, such as potential free-riding or conflicts, might overwhelm the benefits (faster trust formation and easier communication) when the specific dimension of social similarity (such as caste) is less relevant to the task at hand, that is, building a valuable venture. Nevertheless, the finding that caste similarity is not a significant driver of VC funding in modern sectors of the Indian economy seems like good news for Indian policymakers, since enlarging the set of potential economic exchange partners has positive welfare implications. Finally, as outlined below, we conjecture that the implications of our findings are relevant for broader organizational science research.

Some implications for broader organizational science research

Our nuanced conceptualization of the effects of social attributes on outcomes in venture capital markets in an emerging economy is an instance of the globalization of organization theory that has expanded out of its original Western context into more international and often emerging economy contexts. From this vantage point, we offer some conjectures of potential research paths forward to examine relevant issues in organizational settings.

First, our model could usefully connect to the broad research on inequality (Bidwell et al. 2013; Piketty 2014) in mature economies. Specifically, we speculate that issues of immigrant versus non-immigrant employees within new ventures, VC funding of immigrant versus non-immigrant led ventures and deal flow stratification for immigrant versus non-immigrant led VC firms offer fruitful contexts to test and extend our conceptual model.

Second, to the extent entrepreneurship and venture financing are processes of learning and experimentation, our model suggests interesting predictions on misattributions after outcomes are known. VCs likely may be more lenient towards entrepreneurs that are socially similar when attributing responsibility for failures during the process of experimentation. Conversely, dissimilar entrepreneurs may suffer an attribution penalty when successful outcomes are ascribed more to luck than the entrepreneurs' effort. Likewise, we would expect less active monitoring and more advice and counsel (Garg 2013) from VCs who sit on the board of portfolio companies led by socially similar founders. Third, our model is applicable to partner selection and resource exchange under high uncertainty between actors with unequal power. Hence, our model could potentially inform economic exchange in related settings, such as syndicate formation in emerging markets between local and Western VCs, in reverse mergers where firms in an emerging economy list on Western stock markets (Naumovska and Zajac 2017) as well as perhaps in contiguous settings, such as international joint ventures between companies in an emerging economy and Western multinational firms. Finally, it would be worthwhile to investigate whether our results generalize to VC funding outcomes in other emerging economies across Asia (such as Indonesia or Bangladesh) and Latin America (such as Brazil), which are significantly diverse and slowly acquiring the building blocks of an entrepreneurial ecosystem. Because these societies are at different stages of economic development, we should expect interesting differences in how VCs and founders match in these countries. Overall, our model helps further explore how social structure impacts markets for entrepreneurial finance in emerging economies.

REFERENCES

Amit, R., J. Brander, C. Zott. 1998. Why do venture capital firms exist? Theory and Canadian evidence. *Journal of business Venturing* **13**(6) 441-466.

Amit, R., L. Glosten, E. Muller. 1990. Entrepreneurial ability, venture investments, and risk sharing. *Management Science* **36**(10) 1233-1246.

Bapuji, H., S. Chrispal, B. Vissa, G. Ertug. 2019. Implications of Caste for MNEs and IB Research *AIB Annual Meeting*, Copenhagen.

Bengtsson, O., D.H. Hsu. 2015. Ethnic matching in the U.S. venture capital market. *Journal of Business Venturing* **30**(2) 338-354.

Bengtsson, O., B.A. Sensoy. 2011. Investor abilities and financial contracting: Evidence from venture capital. *Journal of Financial Intermediation* **20**(4) 477-502.

Berger, J., S.J. Rosenholtz, M. Zelditch, Jr. 1980. Status Organizing Processes. *Annual Review of Sociology* **6** 479-508.

Bhagavatula, S., M. Bhalla, M. Goel, B. Vissa. 2019. Cultural Diversity in Corporate Boards and Firm Outcomes. INSEAD Working Paper Series.

Bhalla, M., M. Goel, V.T. Konduri, M. Zemel. 2019. Firms of a Feather Merge Together: Cultural Proximity and M&A Outcomes.

Bidwell, M., F. Briscoe, I. Fernandez-Mateo, A. Sterling. 2013. The employment relationship and inequality: How and why changes in employment practices are reshaping rewards in organizations. *Acad. Manag. Ann.* **7**(1) 61-121.

Bonacich, P. 1987. Power and Centrality: A Family of Measures. *American Journal of Sociology* **92** 1170-1182.

Brewer, M.B. 1979. In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological bulletin* **86**(2) 307.

Buchner, A., A. Mohamed, A. Schwienbacher. 2017. Diversification, risk, and returns in venture capital. *Journal of Business Venturing* **32**(5) 519-535.

Cable, D.M., S. Shane. 1997. A prisoner's dilemma approach to entrepreneur-venture capitalist relationships. *Academy of Management Review* **22**(1) 142-176.

Cameron, A.C., J.B. Gelbach, D.L. Miller. 2011. Robust inference with multiway clustering. *Journal of Business & Economic Statistics* **29**(2).

Castellucci, F., G. Ertug. 2010. What's in it for Them? Advantages of Higher-Status Partners in Exchange Relationships. *Academy of Management Journal* **53**(1) 149-166.

Chen, G., R. Chittoor, B. Vissa. 2015. Modernizing without Westernizing: Social Structure and Economic Action in the Indian Financial Sector. *Academy of Management Journal* **58**(2) 26.

Clough, D.R., T.P. Fang, B. Vissa, A. Wu. 2019. Turning Lead Into Gold: How do Entrepreneurs Mobilize Resources to Exploit Opportunities? *Acad. Manag. Ann.* **13**(1) 240-271.

Cochrane, J.H. 2005. The risk and return of venture capital. *Journal of financial economics* **75**(1) 3-52.

Damaraju, N.L., A.K. Makhija. 2018. The role of social proximity in professional CEO appointments: Evidence from caste/religion-based hiring of CEOs in India. *Strategic Management Journal* **39**(7) 2051-2074.

Dimov, D., P.M. de Holan. 2010. Firm Experience and Market Entry by Venture Capital Firms (1962-2004). *Journal of Management Studies* **47**(1) 130-161.

Dimov, D., H. Milanov. 2010. The interplay of need and opportunity in venture capital investment syndication. *Journal of Business Venturing* **25**(4) 331-348.

Drover, W., L. Busenitz, S. Matusik, D. Townsend, A. Anglin, G. Dushnitsky. 2017. A Review and Road Map of Entrepreneurial Equity Financing Research: Venture Capital, Corporate Venture Capital, Angel Investment, Crowdfunding, and Accelerators. *Journal of Management* 0149206317690584.

Dumont, L. 1980. *Homo hierarchicus: the caste system and its implications*. University of Chicago Press.

Eagly, A.H., W. Wood. 1982. Inferred sex differences in status as a determinant of gender stereotypes about social influence. *Journal of personality and social psychology* **43**(5) 915.

Forbes, D.P., M.A. Korsgaard, H.J. Sapienza. 2010. Financing decisions as a source of conflict in venture boards. *Journal of Business Venturing* **25**(6) 579-592.

Franke, N., M. Gruber, D. Harhoff, J. Henkel. 2006. What you are is what you like—similarity biases in venture capitalists' evaluations of start-up teams. *Journal of Business Venturing* **21**(6) 802-826.
Gabarro, J.J. 1990. The development of working relationships. *Intellectual teamwork: Social and technological foundations of cooperative work* 79-110.
Gale, D., L.S. Shapley. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* **69**(1) 9-15.
Garg, S. 2013. Venture boards: Distinctive monitoring and implications for firm performance. *Academy of Management Review* **38**(1) 90-108.
Ge, D., J.M. Mahoney, J.T. Mahoney. 2005. New venture valuation by venture capitalists: An integrative approach. *University of Illinois at Urban Champaign Working Paper* **124** 05-0124.
Geertz, C. 1963. *Peddlers and princes: Social development and economic change in two Indonesian towns*. University of Chicago Press.
Ghurye, G.S. 1969. *Caste and race in India*. Popular Prakashan.
Gompers, P., A. Kovner, J. Lerner, D. Scharfstein. 2008. Venture capital investment cycles: The impact of public markets. *Journal of Financial Economics* **87**(1) 1-23.
Gompers, P., J. Lerner. 2000. Money chasing deals? The impact of fund inflows on private equity valuation. *Journal of Financial Economics* **55**(2) 281-325.
Gompers, P., J. Lerner. 2001. The venture capital revolution. *Journal of economic perspectives* 145-168.
Gornall, W., I.A. Strebulaev. 2017. Squaring Venture Capital Valuations with Reality.
Granovetter, M. 1993. The nature of economic relationships. *Explorations in economic sociology* 3.
Gulati, R., M.C. Higgins. 2003. Which ties matter when? The contingent effects of interorganizational partnerships on IPO success. *Strategic Management Journal* **24**(2) 127-144.
Hallen, B.L. 2008. The Causes and Consequences of the Initial Network Positions of New Organizations: From Whom Do Entrepreneurs Receive Investments? *Administrative Science Quarterly* **53**(4) 685-718.
Hallen, B.L., K.M. Eisenhardt. 2012. Catalyzing strategies and efficient tie formation: how entrepreneurial firms obtain investment ties. *Academy of Management Journal* **55**(1) 35-70.
Hart, O. 2001. Financial Contracting. *Journal of Economic Literature* **39**(4) 1079-1100.
Hawawini, G., C. Viallet. 2010. *Finance for executives: Managing for value creation*. Cengage Learning.
Heckman, J.J. 1979. Sample selection bias as a specification error. *Econometrica* **47** 153-161.
Hegde, D., J. Tumlinson. 2014. Does Social Proximity Enhance Business Partnerships? Theory and Evidence from Ethnicity's Role in US Venture Capital. *Management Science* **60**(9) 2355-2380.
Hellmann, T., M. Puri. 2002. Venture capital and the professionalization of start-up firms: Empirical evidence. *Journal of Finance* **57**(1) 169-197.
Hochberg, Y.V., A. Ljungqvist, Y. Lu. 2007. Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance* **62**(1) 251-301.
Hochberg, Y.V., A. Ljungqvist, Y. Lu. 2010. Networking as a Barrier to Entry and the Competitive Supply of Venture Capital. *Journal of Finance* **65**(3) 829-859.
Hölmstrom, B. 1979. Moral hazard and observability. *The Bell journal of economics* 74-91.
Homans, G.C. 1950. The human group new york. *Harpers*.
Hsu, D.H. 2004. What Do Entrepreneurs Pay for Venture Capital Affiliation? *The Journal of Finance* **59**(4) 1805-1844.
Hsu, D.H., R.H. Ziedonis. 2013. Resources as dual sources of advantage: Implications for valuing entrepreneurial-firm patents. *Strategic Management Journal* **34**(7) 761-781.
Huang, L., A.P. Knight. 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.
Kachru, B.B. 1983. *The indianization of English: the English language in India*. Oxford University Press Oxford.
Kaplan, S.N., P. Strömberg. 2003. Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts. *The Review of Economic Studies* **70**(2) 281-315.
Keay, J. 2011. *India: A History. Revised and Updated*. Grove/Atlantic, Inc.
Kerr, W.R., R. Nanda, M. Rhodes-Kropf. 2014. Entrepreneurship as experimentation. *The Journal of Economic Perspectives* **28**(3) 25-48.

Klein, T. 2016. Analysis of Stable Matchings in R: Package matchingMarkets. *The Comprehensive R Archive Network.*
Krishnamurti, B. 1990. The regional language vis-à-vis English as the medium of instruction in higher education: the Indian dilemma. *Multilingualism in India* 15-24.
Larson, A. 1992. Network dyads in entrepreneurial settings: A study of the governance of exchange relationships. *Administrative science quarterly* 76-104.
Ljungqvist, A., M. Richardson. 2003. The cash flow, return and risk characteristics of private equity. National Bureau of Economic Research.
Manigart, S., K. De Waele, M. Wright, K. Robbie, P. Desbrières, H.J. Sapienza, A. Beekman. 2002. Determinants of required return in venture capital investments: a five-country study. *Journal of Business Venturing* **17**(4) 291-312.
Matusik, S.F., M.A. Fitza. 2012. Diversification in the venture capital industry: leveraging knowledge under uncertainty. *Strategic Management Journal* **33**(4) 407-426.
McPherson, M., L. Smith-Lovin, J.M. Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* **27** 415-444.
Menning, G. 1997. Trust, entrepreneurship and development in Surat city, India*. *Ethnos* **62**(1-2) 59-90.
Merton, R.K. 1948. The self-fulfilling prophecy. *The Antioch Review* **8**(2) 193-210.
Mitsuhashi, H., H.R. Greve. 2004. Powerful and free: Intraorganizational power and the dynamics of corporate strategy. *Strategic Organization* **2**(2) 107-132.
Nahata, R. 2008. Venture capital reputation and investment performance. *Journal of Financial Economics* **90**(2) 127-151.
Naumovska, I., E.J. Zajac. 2017. Who and/or What Gets Blamed? How U.S. Financial Markets Penalize Chinese Firms. INSEAD Working Paper
Ozmel, U., J. Reuer, R. Gulati. 2013. Signals across multiple networks: How venture capital and alliance networks affect interorganizational collaboration. *Academy of Management Journal* **53**(3) 14.
Pahnke, E., R. McDonald, D. Wang, B. Hallen. 2014. Exposed: Venture Capital, Competitor Ties, and Entrepreneurial Innovation. *Academy of Management Journal* amj. 2012.0777.
Petersen, M.A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of financial studies* **22**(1) 435-480.
Piketty, T. 2014. *Capital in the twenty-first century*. Harvard University Press.
Ramaswamy, S. 1997. *Passions of the tongue: Language devotion in Tamil India, 1891-1970*. Univ of California Press.
Ridgeway, C. 1991. The Social Construction of Status Value: Gender and Other Nominal Characteristics. *Social Forces* **70**(2) 367-386.
Rogers, E.M., D.K. Bhowmik. 1970. Homophily-heterophily: Relational concepts for communication research. *Public opinion quarterly* **34**(4) 523-538.
Sahlman, W.A. 1990. The structure and governance of venture-capital organizations. *Journal of financial economics* **27**(2) 473-521.
Sanders, W.M.G., S. Boivie. 2004. Sorting Things Out: Valuation of New Firms in Uncertain Markets. *Strategic Management Journal* **25**(2) 167-186.
Sauder, M., F. Lynn, J.M. Podolny. 2012. Status: Insights from organizational sociology. *Annual Review of Sociology* **38** 267-283.
Shipilov, A.V., S.X. Li, H.R. Greve. 2011. The prince and the pauper: Search and brokerage in the initiation of status-heterophilous ties. *Organization Science* **22**(6) 1418-1434.
Sørensen, M. 2007. How smart is smart money? A two-sided matching model of Venture Capital. *The Journal of Finance* **62**(6) 2725-2762.
Sorenson, O., T.E. Stuart. 2001. Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology* **106**(6) 1546-1588.
Sorenson, O., D.M. Waguespack. 2006. Social structure and exchange: Self-confirming dynamics in Hollywood. *Administrative Science Quarterly* **51**(4) 560-589.
Srinivas, M.N. 1957. Caste in modern India. *The Journal of Asian Studies* 529-548.

Sunil, K. 1997. The idea of India. *Hamish Hamilton London*.
Tenbrunsel, A.E., K.A. Wade-Benzoni, J. Moag, M.H. Bazerman. 1999. The negotiation matching process: Relationships and partner selection. *Organizational Behavior and Human Decision Processes* **80**(3) 252-283.
Vissa, B. 2011. A matching theory of entrepreneurs' tie formation intentions and initiation of economic exchange. *Academy of Management Journal* **54**(1) 137-158.
Vissa, B. 2012. Agency in action: Entrepreneurs' networking style and initiation of economic exchange. *Organization Science* **23**(2) 492-510.
Wiser, W.H. 1988. The Hindu jajmani system.
Yenkey, C.B. 2017. Fraud and Market Participation. *Administrative Science Quarterly* **0**(0) 0001839217694359.
Zacharakis, A.L., D.A. Shepherd. 2001. The nature of information and overconfidence on venture capitalists' decision making. *Journal of Business Venturing* **16**(4) 311-332.
Zhang, J., P.K. Wong, Y.P. Ho. 2016. Ethnic enclave and entrepreneurial financing: Asian venture capitalists in Silicon Valley. *Strategic Entrepreneurship Journal* **10**(3) 318-335.
Zucker, L.G. 1986. Production of trust: Institutional sources of economic structure, 1840–1920. *Research in organizational behavior*.

Figure 1: Construction of the Caste Superiority Index

Deal #X and Deal #Y below involve two founders (E1, E2) and two VCs (VC1, VC2). There are four possible entrepreneur–VC dyads per deal (column A).

The overall caste "value" of an entrepreneur (column I) is calculated as the sum product of this entrepreneur's caste distribution frequencies (columns D to H) and the ratio of each caste to the untouchable caste (first line of the table). These ratios show how much more a given caste is naturally represented relative to the untouchable caste in our setting. We calculate the caste value for a VC (column O) similarly using the caste ratios and the frequencies in columns J to N.

We compute the caste value for each entrepreneur and each VC involved in the deal (columns I and O). Then we calculate the difference between these values for each dyad (column P).

Caste superiority index is the average of these differences at the deal level (i.e., the average of column P).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
		caste ratio		3.9	5.28	3.08	3.1	1		3.9	5.28	3.08	3.1	1		
				Entrepreneur caste frequency						**VC caste frequency**						
Deal	**dyads**	**Ent. name**	**VC. name**	priest	warrior	merch.	peasant	untouch.	**value**	priest	warrior	merch.	peasant	untouch.	**value**	**Δ**
#X	E1-VC1	Menon	Navani	4.8%	93.0%	0.3%	1.6%	0.2%	5.2	0.0%	9.1%	72.7%	18.2%	0.0%	3.3	**1.9**
	E1-VC2	Menon	Bajaj	4.8%	93.0%	0.3%	1.6%	0.2%	5.2	2.5%	24.8%	70.5%	1.6%	0.6%	3.6	**1.5**
	E2-VC1	Nautiyal	Navani	92.4%	5.6%	0.0%	0.4%	1.6%	3.9	0.0%	9.1%	72.7%	18.2%	0.0%	3.3	**0.6**
	E2-VC2	Nautiyal	Bajaj	92.4%	5.6%	0.0%	0.4%	1.6%	3.9	2.5%	24.8%	70.5%	1.6%	0.6%	3.6	**0.3**
#Y	E1-VC1	Patel	Trivedi	1.7%	2.1%	81.1%	11.9%	3.2%	3.1	99.1%	0.2%	0.2%	0.3%	0.1%	3.9	**-0.8**
	E1-VC2	Patel	Bakshi	1.7%	2.1%	81.1%	11.9%	3.2%	3.1	84.5%	8.2%	3.1%	2.6%	1.6%	3.9	**-0.8**
	E2-VC1	Soni	Trivedi	4.7%	10.6%	7.1%	74.7%	2.8%	3.3	99.1%	0.2%	0.2%	0.3%	0.1%	3.9	**-0.6**
	E2-VC2	Soni	Bakshi	4.7%	10.6%	7.1%	74.7%	2.8%	3.3	84.5%	8.2%	3.1%	2.6%	1.6%	3.9	**-0.6**

Caste superiority index (Deal #X) = (1.9 + 1.5 + 0.6 + 0.3) / 4 **= 1.1**

Caste superiority index (Deal #Y) = (-0.8 -0.8 - 0.6 - 0.6) / 4 **= -0.7**

For Deal #X, caste superiority index is greater than 0.05 therefore we consider that the founders are on average of higher caste than the VCs in this deal. It is the reverse in Deal #Y.

Table 1a

Heckman Second-Stage: Valuation Variables – Means, Standard Deviations, and Correlations

Variables	Mean	s.d.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Valuation (log)	5.55	1.41																
2. Inverse Mills ratio	0.90	0.59	.06															
3. Number of prior funding events	0.51	0.84	.32	-.28														
4. Early stage	0.66	0.48	-.52	-.15	-.19													
5. Venture age	4.21	3.57	.24	.07	.19	-.43												
6. Founding team eliteness	1.14	1.30	-.12	-.30	.08	.24	-.14											
7. Start-up experience	0.08	0.52	.14	-.03	-.02	-.08	-.03	-.01										
8. Proportion of merchant caste	0.23	0.37	.11	.06	-.03	-.02	-.04	-.02	.10									
9. Number of board seats	0.71	0.68	-.07	.11	-.22	.08	-.13	.04	-.03	-.01								
10. VC firm status	0.21	0.35	.15	-.20	.21	.07	-.04	.11	-.01	.00	.03							
11. Entering new industry	0.65	0.67	.06	.03	-.03	-.10	.02	-.05	-.05	-.01	.01	.05						
12. VC experience	2.89	3.63	.01	-.18	.14	.09	-.09	.09	.01	-.03	.06	.49	.04					
13. VC competitive intensity	7.72	6.68	-.02	.08	-.09	.03	-.04	-.02	.10	.17	.00	-.03	-.05	.07				
14. Industry hotness	669.8	516.92	.02	.06	-.04	-.03	.04	-.07	.04	.05	-.05	-.08	.12	-.04	-.06			
15. Geographical distance	0.41	0.55	.05	.44	-.08	-.10	.10	-.02	-.06	.03	.00	-.07	.04	-.09	-.19	.04		
16. Caste similarity	-1.17	0.99	.00	-.26	.11	.06	-.02	.21	.07	.08	.12	.18	-.02	.14	-.02	-.03	-.18	
17. Regional similarity	-2.41	1.95	.08	-.18	.11	.09	-.12	.13	-.01	.05	.14	.17	-.02	.09	-.02	-.06	-.17	.49

Correlations ≥ |.08| are significant at $p \leq .05$; $n = 622$. For our independent variables of interest and moderators we report the mean and standard deviation before mean-centering for ease of interpretation.

Table 1b

Heckman Second-Stage: Downside Risk Protection Variables - Means, Standard Deviations, and Correlations

	Variables	Mean	s.d.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1.	Downside risk protection	4.67	3.37																
2.	Inverse Mills ratio	0.90	0.59	-.06															
3.	Number of prior funding events	0.50	0.84	-.02	-.26														
4.	Early stage	0.67	0.47	.07	-.17	-.19													
5.	Venture age	4.15	3.57	.03	.08	.20	-.42												
6.	Founding team eliteness	1.16	1.31	.08	-.31	.08	.22	-.13											
7.	Start-up experience	0.08	0.50	-.03	-.01	-.03	-.07	-.02	.00										
8.	Proportion of merchant caste	0.22	0.37	-.01	.06	-.05	.00	-.04	.01	.08									
9.	Number of board seats	0.73	0.68	.09	.10	-.22	.06	-.11	.04	-.02	.01								
10.	VC firm status	0.21	0.35	.10	-.21	.22	.07	-.03	.11	-.01	.00	.03							
11.	Entering new industry	0.65	0.67	-.02	.01	-.01	-.10	.01	-.04	-.05	.01	-.01	.05						
12.	VC experience	2.90	3.62	.10	-.20	.15	.09	-.07	.08	.02	-.02	.06	.49	.05					
13.	VC competitive intensity	7.77	6.71	-.02	.10	-.11	.03	-.04	-.02	.08	.17	.00	-.03	-.04	.07				
14.	Industry hotness	673.6	525.09	-.06	.06	-.04	-.04	.04	-.08	.03	.06	-.05	-.08	.12	-.05	-.05			
15.	Geographical distance	0.41	0.55	-.14	.43	-.07	-.11	.09	-.01	-.05	.02	.01	-.07	.04	-.10	-.18	.04		
16.	Caste similarity	-1.17	0.99	.14	-.25	.09	.09	-.03	.22	.07	.07	.14	.19	-.01	.14	-.03	-.03	-.17	
17.	Regional similarity	-2.38	1.85	.17	-.18	.10	.09	-.11	.14	.02	.04	.15	.17	-.01	.10	-.04	-.05	-.14	.52

Correlations ≥ |.08| are significant at $p \leq .05$; $n = 590$. For our independent variables of interest and moderators we report the mean and standard deviation before mean-centering for ease of interpretation.

Table 2: Heckman First-Stage Variables - Means, Standard Deviations, and Correlations

Variables		Mean	s.d.	1	2	3	4	5	6	7	8	9	10	11
1.	Funded	0.22	0.42											
2.	Number of prior funding events	0.83	1.03	-.17										
3.	Founding team eliteness	0.77	1.02	.16	.04									
4.	Start-up experience	0.08	0.51	.02	-.06	-.01								
5.	Proportion of merchant caste	0.24	0.38	-.02	-.02	-.01	.08							
6.	Geographical distance	0.70	0.58	-.27	.01	-.04	.00	-.01						
7.	New tie	0.94	0.23	-.41	-.19	-.12	.00	.01	.17					
8.	Industry hotness	683.77	611.27	.00	-.06	-.03	.05	.02	.05	.01				
9.	Caste similarity	-1.35	0.99	.09	.07	.13	.00	.02	-.09	-.12	.00			
10.	Regional similarity	-2.76	1.80	.09	.02	.10	.03	.04	-.10	-.12	-.05	.43		
11.	Median caste similarity	-0.99	0.66	.04	-.12	-.07	-.03	-.07	.01	.05	.08	.16	.18	
12.	Median regional similarity	-2.28	1.28	-.03	.04	.04	-.04	.05	-.05	-.04	.00	.06	.35	.41

Correlations ≥ |.04| are significant at $p \leq .05$; n = 3,285. For our independent variables of interest we report the mean and standard deviation before mean-centering for ease of interpretation.

Table 3: Financial Returns - Means, standard deviations, and correlations

Variables	Mean	s.d.	1	2	3	4	5	6	7	8	9	10	11	12	13
1. IRR (log)	2.41	2.48													
2.VC firm status	0.10	0.20	.11												
3.Founding team eliteness	0.88	1.26	.13	-.01											
4.Proportion of merchant caste	0.21	0.38	-.04	-.05	-.01										
5.Total amount	215.92	222.63	.17	.31	-.04	.12									
6.First investment valuation	762.50	849.51	.18	.2	-.21	.07	.57								
7.Downside risk protection	4.70	3.73	-.06	.16	.09	-.04	-.06	-.08							
8.Number of rounds	1.24	0.71	.35	-.01	.31	.08	.22	-.07	-.09						
9.Number of ventures	0.90	1.26	.02	.60	-.03	.08	.24	.04	.20	.10					
10.Exit year	2011.4	2.09	.07	.13	.13	-.01	.26	.02	.00	.22	.33				
11.Geographical distance	0.40	0.50	-.01	.08	-.03	.01	.10	.18	-.14	-.1	-.12	.09			
12.IT	0.35	0.48	-.14	-.01	.20	.19	-.21	-.30	.08	.08	.11	.00	-.03		
13.Caste similarity	-1.43	1.17	.07	.12	.23	.01	.03	-.11	.09	.20	.21	.15	-.05	-.02	
14.Regional similarity	-2.62	1.92	.15	.21	.10	.01	.01	-.13	.11	.08	.13	.04	-.07	.02	.52

Correlations ≥ |.16| are significant at $p \leq .05$; n = 145. For our independent variables of interest we report the mean and standard deviation before mean-centering for ease of interpretation.

Table 4: Probit Estimates for Heckman First-Stage - Receiving Funding

Variable	Funded
Number of prior funding events	**-0.658*****
	(0.044)
Founding team eliteness	**0.194*****
	(0.028)
Start-up experience	**0.008**
	(0.052)
Proportion of merchant caste	**-0.043**
	(0.079)
Geographical distance	**-0.726*****
	(0.057)
New tie	**-3.018*****
	(0.172)
Industry hotness	**1.4E-05**
	(4.6E-05)
Caste similarity	**0.010**
	(0.033)
Regional similarity	**0.033***
	(0.020)
Median caste similarity	**-0.074+**
	(0.049)
Median regional similarity	**-0.078****
	(0.028)
Region dummies	Yes
Investor type dummies	Yes
Constant	**2.738*****
	(0.207)
Pseudo R-Square	0.31
Log-likelihood	-1204.09
Chi-square	1069.48
n	3285

N=3285, one-tailed tests, estimated coefficients are in bold, standard errors are in parentheses. The regression is significant ($p < 0.0001$). *** p-value<0.001, ** p value<0.01, *p-value<0.05, + p-value<0.1

Table 5: Panel A - OLS Regressions of Financial Returns of VC Firms (IRR) Panel B – Two-Sided Matching Model and Bayesian Estimates of Financial Returns of VC Firms

Variable	IRR Panel A		IRR Panel B	
	Model 1	Model 2	Model 3 (Bayesian)	Model 4 (OLS)
VC firm status	**2.064***	**1.458**	**0.984**	**1.617**
	(1.033)	(1.252)	(1.207)	(1.259)
Founding team eliteness	**0.095**	**0.122**	**0.096**	**0.085**
	(0.163)	(0.163)	(0.177)	(0.167)
Proportion of merchant caste	**0.034**	**-0.020**	**-0.005**	**0.001**
	(0.513)	(0.540)	(0.553)	(0.545)
Total amount	**-0.001**	**-0.001**	**-0.001**	**-0.001**
	(0.001)	(0.001)	(0.001)	(0.001)
First investment valuation	**4.9E-04*****	**0.001***	**0.001***	**0.001***
	(1.4E-04)	(0.000)	(3.3E-04)	(0.000)
Downside risk protection	**-0.014**	**-0.018**	**-0.044**	**-0.039**
	(0.054)	(0.053)	(0.055)	(0.056)
Number of rounds	**1.336*****	**1.381*****	**1.4*****	**1.319*****
	(0.260)	(0.262)	(0.332)	(0.262)
Number of ventures	**-0.257**	**-0.174**	**-0.054**	**-0.187**
	(0.248)	(0.255)	(0.208)	(0.263)
Exit year	**0.020**	**0.015**	**3.5E-04**	**0.031**
	(0.107)	(0.114)	(0.002)	(0.120)
Geographical distance	**-0.155**	**-0.126**	**0.053**	**-0.014**
	(0.404)	(0.422)	(0.423)	(0.430)
IT	**-0.731+**	**-0.737+**	**-0.705+**	**-0.735+**
	(0.430)	(0.454)	(0.462)	(0.465)
Caste similarity	**-**	**-0.256**	**-0.329+**	**-0.280+**
	-	(0.204)	(0.205)	(0.212)
Regional similarity	**-**	**0.221***	**0.244***	**0.211+**
	-	(0.111)	(0.123)	(0.118)
Region dummies	Yes	Yes	Yes	Yes
Constant	**-39.608**	**-30.456**	**0.072**	**-61.056**
	(214.880)	(229.086)	(3.183)	(240.507)
	N/A	N/A	**2-sided matching**	N/A
VC firm status			**-0.347**	
			(0.409)	
Founding team eliteness			**-0.128**	
			(0.172)	
Proportion of merchant caste			**-1.103+**	
			(0.686)	

Geographical distance			**1.504***	
			(0.77)	
IT			**-1.243***	
			(0.604)	
Median caste similarity			**-1.587***	
			(0.679)	
Median regional similarity			**-0.943****	
			(0.329)	
Caste similarity			**1.028****	
			(0.394)	
Regional similarity			**-0.594****	
			(0.201)	
Region dummies			Yes	
Investor type dummies			Yes	
F-Statistic	9.69	7.17	-	6.28
R-Square	0.248	0.251	-	0.237
N	148	145	138	138
	N/A	N/A	**Variance**	N/A
Covariance			0.164	
			(0.156)	

one-tailed tests, estimated coefficients are in bold, standard errors are in parentheses. The regressions are significant (p-value<0.0001across models). *** p-value<0.001, ** p-value<0.01, * p-value<0.05, + p-value<0.1

Table 6: Estimates for Heckman Second-Stage Models - Drivers of Venture Valuation

Variables	Valuation				
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Inverse Mills ratio	**0.058**	**0.062**	**0.048**	**0.067**	**0.053**
	(0.108)	(0.109)	(0.113)	(0.112)	(0.116)
Number of prior funding events	**0.358*****	**0.341*****	**0.344*****	**0.339*****	**0.341*****
	(0.077)	(0.072)	(0.072)	(0.074)	(0.073)
Early stage	**-1.227*****	**-1.243*****	**-1.241*****	**-1.251*****	**-1.248*****
	(0.134)	(0.137)	(0.132)	(0.137)	(0.132)
Venture age	**0.002**	**0.006**	**0.004**	**0.004**	**0.002**
	(0.018)	(0.018)	(0.018)	(0.017)	(0.017)
Founding team eliteness	**-0.037**	**-0.027**	**-0.033**	**-0.025**	**-0.031**
	(0.042)	(0.041)	(0.042)	(0.041)	(0.042)
Start-up experience	**0.206****	**0.212****	**0.205****	**0.202****	**0.196****
	(0.070)	(0.072)	(0.074)	(0.074)	(0.075)
Proportion of merchant caste	**0.407****	**0.402****	**0.401****	**0.407****	**0.405****
	(0.142)	(0.147)	(0.146)	(0.144)	(0.143)
Number of board seats	**0.043**	**0.024**	**0.025**	**0.031**	**0.031**
	(0.081)	(0.082)	(0.081)	(0.080)	(0.080)
VC firm status	**0.765*****	**0.702*****	**0.693*****	**0.712*****	**0.702*****
	(0.190)	(0.185)	(0.184)	(0.179)	(0.178)
Entering new industry	**-0.007**	**-0.010**	**0.006**	**-0.022**	**-0.007**
	(0.082)	(0.079)	(0.078)	(0.075)	(0.075)
VC experience	**-0.022+**	**-0.019+**	**-0.020+**	**-0.023***	**-0.023***
	(0.014)	(0.013)	(0.013)	(0.012)	(0.013)
VC competitive intensity	**-0.003**	**-0.004**	**-0.005**	**-0.003**	**-0.004**
	(0.016)	(0.015)	(0.014)	(0.014)	(0.014)
Industry hotness	**-1.5E-04**	**-1.7E-04**	**-1.6E-04**	**-1.9E-04**	**-1.8E-04**
	(1.8E-04)	(1.8E-04)	(1.8E-04)	(1.9E-04)	(1.9E-04)
Geographical distance	**0.016**	**0.011**	**0.004**	**0.002**	**-0.004**
	(0.104)	(0.106)	(0.107)	(0.107)	(0.107)
Caste similarity	**-**	**-0.117***	**-0.125***	**-0.114***	**-0.122***
	-	(0.058)	(0.057)	(0.056)	(0.056)
Regional similarity	-	**0.110*****	**0.118*****	**0.108****	**0.114*****
	-	(0.035)	(0.033)	(0.033)	(0.032)
Early stage × Caste similarity	-	**-**	**0.033**	**-**	**0.028**
	-	**-**	(0.144)	**-**	(0.141)
Early stage × Regional similarity	-	**-**	**0.122+**	**-**	**0.110+**
	-	**-**	(0.080)	**-**	(0.075)
Entering new industry × Caste similarity	**-**	**-**	**-**	**-0.303*****	**-0.283*****
	-	-	-	(0.085)	(0.085)
Entering new industry × Regional similarity	-	-	-	**0.149****	**0.145****
	-	-	-	(0.052)	(0.052)
Industry dummies	Yes	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes	Yes

Investor type dummies	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
Constant	**4.585*** **	**5.171*** **	**5.183*** **	**4.684*** **	**4.711*** **
	(0.629)	(0.642)	(0.688)	(0.659)	(0.702)
F-Statistic	10.534	11.048	10.971	11.777	17.781
Adjusted R-Square	0.420	0.429	0.433	0.439	0.442

n = 622; one-tailed tests, estimated coefficients are in bold, standard errors clustered for ventures and for VC firms are in parentheses. The regressions are significant (p-value<0.0001 across models). *** p-value<0.001, ** p-value<0.01, * p-value<0.05, + p-value<0.1

Table 7: Estimates for Heckman Second-Stage Models - Drivers of Downside Risk Protection

Variables	**Downside Risk Protection**				
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Inverse Mills ratio	**0.265**	**0.302**	**0.290**	**0.300**	**0.284**
	(0.295)	(0.295)	(0.298)	(0.296)	(0.298)
Number of prior funding events	**-0.227**	**-0.273+**	**-0.273+**	**-0.270+**	**-0.270+**
	(0.185)	(0.185)	(0.186)	(0.186)	(0.186)
Early stage	**0.236**	**0.217**	**0.220**	**0.263**	**0.269**
	(0.347)	(0.345)	(0.346)	(0.350)	(0.350)
Venture age	**0.089***	**0.093***	**0.092***	**0.095***	**0.094***
	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Founding team eliteness	**0.197***	**0.180+**	**0.177+**	**0.175+**	**0.170+**
	(0.113)	(0.114)	(0.114)	(0.114)	(0.115)
Start-up experience	**-0.064**	**-0.104**	**-0.104**	**-0.104**	**-0.103**
	(0.290)	(0.290)	(0.291)	(0.291)	(0.292)
Proportion of merchant caste	**-0.106**	**-0.163**	**-0.168**	**-0.178**	**-0.185**
	(0.385)	(0.385)	(0.385)	(0.386)	(0.386)
Number of board seats	**0.261**	**0.175**	**0.176**	**0.170**	**0.170**
	(0.215)	(0.217)	(0.218)	(0.218)	(0.218)
VC firm status	**0.913***	**0.704***	**0.698+**	**0.691+**	**0.683+**
	(0.461)	(0.467)	(0.467)	(0.467)	(0.468)
Entering new industry	**0.205**	**0.182**	**0.195**	**0.170**	**0.186**
	(0.217)	(0.216)	(0.218)	(0.217)	(0.219)
VC experience	**0.040**	**0.044**	**0.042**	**0.048**	**0.047**
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
VC competitive intensity	**-0.032+**	**-0.028+**	**-0.029+**	**-0.027**	**-0.028**
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Industry hotness	**2.7e-04**	**1.9e-04**	**1.9e-04**	**2.1e-04**	**2.1e-04**
	(3.8e-04)	(3.8e-04)	(3.8e-04)	(3.8e-04)	(3.8e-04)
Geographical distance	**-0.849****	**-0.802****	**-0.803****	**-0.788****	**-0.790****
	(0.291)	(0.292)	(0.292)	(0.292)	(0.293)
Caste similarity	**-**	**0.037**	**0.032**	**0.041**	**0.036**
	-	(0.171)	(0.171)	(0.171)	(0.171)
Regional similarity	**-**	**0.206***	**0.211***	**0.204***	**0.209***
	-	(0.099)	(0.100)	(0.099)	(0.100)
Early stage × Caste similarity	**-**	**-**	**0.023**	**-**	**0.050**

	-	**-**	(0.352)	**-**	(0.354)
Early stage × Regional similarity	**-**	**-**	**0.082**	**-**	**0.086**
	-	-	(0.196)	-	(0.196)
Entering new industry × Caste similarity	**-**	**-**	**-**	**0.194**	**0.211**
	-	**-**	**-**	(0.289)	(0.291)
Entering new industry × Regional similarity	**-**	**-**	**-**	**0.017**	**0.013**
	-	-	-	(0.153)	(0.153)
Industry dummies	Yes	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes	Yes
Investor type dummies	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
Constant	**-0.098**	**0.504**	**0.582**	**0.589**	**0.680**
	(3.446)	(3.453)	(3.471)	(3.479)	(3.498)
F statistic	2.588	2.631	2.512	2.527	2.420
Adjusted R-Square	0.100	0.106	0.104	0.104	0.102

n = 590; one-tailed tests, estimated coefficients are in bold, robust standard errors are in parentheses. The regressions are significant (p-value<0.0001 across models). *** p-value<0.001, ** p-value<0.01, * p-value<0.05, + p-value<0.1

Table 8: Estimates for Heckman Second-Stage Models - Drivers of Venture Valuation with Caste Dissimilarity Variables

Variable	Valuation					
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Inverse Mills ratio	**0.058**	**0.067**	**0.061**	**0.064**	**0.066**	**0.075**
	(0.108)	(0.111)	(0.113)	(0.114)	(0.114)	(0.111)
Number of prior funding events	**0.358*** **	**0.339*** **	**0.344*** **	**0.333*** **	**0.333*** **	**0.333*** **
	(0.077)	(0.072)	(0.071)	(0.073)	(0.073)	(0.072)
Early stage	**-1.227*** **	**-1.235*** **	**-1.237*** **	**-1.225*** **	**-1.224*** **	**-1.243*** **
	(0.134)	(0.136)	(0.131)	(0.137)	(0.137)	(0.136)
Venture age	**0.002**	**0.006**	**0.004**	**0.005**	**0.005**	**0.005**
	(0.018)	(0.018)	(0.017)	(0.018)	(0.018)	(0.018)
Founding team eliteness	**-0.037**	**-0.030**	**-0.035**	**-0.028**	**-0.028**	**-0.027**
	(0.042)	(0.041)	(0.041)	(0.042)	(0.042)	(0.042)
Start-up experience	**0.206** **	**0.210** **	**0.200** **	**0.206** **	**0.206** **	**0.209** **
	(0.070)	(0.071)	(0.072)	(0.072)	(0.072)	(0.073)
Proportion of merchant caste	**0.407** **	**0.398** **	**0.397** **	**0.405** **	**0.405** **	**0.395** **
	(0.142)	(0.144)	(0.145)	(0.142)	(0.142)	(0.143)
Number of board seats	**0.043**	**0.015**	**0.017**	**0.017**	**0.017**	**0.010**
	(0.081)	(0.080)	(0.078)	(0.079)	(0.079)	(0.079)
VC firm status	**0.765*** **	**0.716*** **	**0.700*** **	**0.731*** **	**0.731*** **	**0.703*** **
	(0.190)	(0.184)	(0.185)	(0.181)	(0.181)	(0.183)
Entering new industry	**-0.007**	**-0.015**	**0.000**	**-0.037**	**-0.037**	**-0.012**
	(0.082)	(0.080)	(0.078)	(0.076)	(0.076)	(0.080)
VC experience	**-0.022**	**-0.018**	**-0.019**	**-0.020**	**-0.020**	**-0.017**
	(0.014)	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)
VC competitive intensity	**-0.003**	**-0.004**	**-0.004**	**-0.004**	**-0.004**	**-0.004**

Variable	Valuation					
	(0.016)	(0.015)	(0.014)	(0.015)	(0.015)	(0.014)
Industry hotness	**-1.5e-04**	**-1.7e-04**	**-1.7e-04**	**-1.8e-04**	**-1.8e-04**	**-1.7e-04**
	(2.0e-04)	(1.8e-04)	(1.8e-04)	(1.9e-04)	(1.8e-04)	(1.8e-04)
Geographical distance	**0.016**	**0.009**	**0.001**	**0.010**	**0.008**	**0.006**
	(0.104)	(0.106)	(0.107)	(0.106)	(0.106)	(0.106)
Caste dissimilarity (VC<ENT)	**-**	**0.031**	**0.045**	**0.031**	**-0.032**	**-0.034**
	-	(0.081)	(0.080)	(0.082)	(0.082)	(0.081)
Caste dissimilarity (VC=ENT)	-	**0.179**	**0.150**	**0.191**	**-0.188**	**-0.185**
	-	(0.170)	(0.164)	(0.170)	(0.170)	(0.170)
Caste dissimilarity (VC>ENT)	-	**0.186****	**0.192****	**0.199****	**0.199****	**0.217****
	-	(0.076)	(0.076)	(0.074)	(0.075)	(0.080)
Regional similarity	-	**0.108*****	**0.116*****	**0.105*****	**0.105*****	**0.111*****
	-	(0.035)	(0.033)	(0.034)	(0.034)	(0.035)
Early stage × Caste dissimilarity (VC>ENT)	-	**-**	**0.124**	**-**	**0.177**	**-**
	-	**-**	(0.166)	**-**	(0.166)	-
Early stage × Regional similarity	-	**-**	**0.144***	**-**	**0.148***	**-**
	-	**-**	(0.074)	**-**	(0.073)	-
Entering new industry × Caste dissimilarity	**-**	**-**	**-**	**0.278***	**0.294***	**-**
(VC>ENT)	-	-	-	(0.131)	(0.133)	-
Entering new industry × Regional similarity	-	-	-	**0.084***	**0.086***	**-**
	-	-	-	(0.045)	(0.045)	-
Founding team eliteness × Caste	-	-	-	-	**-**	**0.097+**
dissimilarity (VC>ENT)	-	-	-	-	-	(0.071)
Founding team eliteness × Regional	-	-	-	**-**	**-**	**0.012**
similarity	-	-	-	-	-	(0.030)
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes
Investor type dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Constant	**4.585*****	**5.230*****	**5.239*****	**4.886*****	**5.186*****	**5.354*****
	(0.629)	(0.638)	(0.723)	(0.660)	(0.637)	(0.641)
F-Statistic	10.534	10.691	10.609	11.053	10.955	10.221
Adjusted R-Square	0.420	0.430	0.434	0.434	0.434	0.430

n = 622; one-tailed tests, estimated coefficients are in bold, robust standard errors are in parentheses. The regressions are significant (p-value<0.0001 across models). *** p-value<0.001, ** p-value<0.01, * p-value<0.05, + p-value<0.1