Junior Management Science

journal homepage: www.jums.academy

Do birds of a feather always flock together? A multidimensional examination of homophily in crowdfunding

Markus Klepsch

Technische Universität München

Abstract

Homophily—the tendency of individuals to associate with similar others—is one of the most persistent findings in research on interpersonal interaction. Literature has recently also studied the impact of homophily in crowdfunding markets (Greenberg & Mollick, 2017). While these results offer valuable insights into the dimension sex, homophily is a multidimensional construct in theory (McPherson, Smith-Lovin, & Cook, 2001). Therefore, I employ homophily theorizing by analyzing to what extent homophily influences backers' choices in crowdfunding regarding five different sociodemographic dimensions. To test my hypotheses, I drew upon a unique dataset of roughly 3.5 million backings made on the crowdfunding platform Indie-gogo. My results suggest that individuals make homophilic choices with respect to age, sex, occupation, and location. While my findings thus confirm that, in large, homophily plays an essential role in individual choices, I extend the literature by showing that it is not always as clear-cut. The more complex the dimension, the more difficult it is to judge. Specifically, I show that in the dimension race, where a plethora of possible combinations leads to a certain degree of complexity, there is no simple 'yes' or 'no' answer to whether homophily influences the decision. Instead, significant differences can be observed depending on the specific race under consideration.

Keywords: Crowdfunding; Indiegogo; Homophily.

1. Introduction

Choice homophily (Greenberg & Mollick, 2015, 2017; McPherson & Smith-Lovin, 1987) is influential in resource acquisition online. In general, the concept of homophily can be described as "like attracted to like" (Greenberg & Mollick, 2017, p. 342) and argues that the basis of attraction between individuals is the similarity between them, where "contact between similar people occurs at a higher rate than among dissimilar people" (McPherson et al., 2001, p. 416). In the past, the concept of homophily has been used to explain a plethora of phenomena in the social sciences ranging from mere personal contact (Wellman, 1996) over friendship (Kandel, 1978; Lazarsfeld & Merton, 1954; Verbrugge, 1977) to the close bond of marriage (Alba & Golden, 1986; Kalmijn, 1998). Essentially, two types of homophily—induced and choice—can be distinguished. While induced homophily describes "structural mechanisms that define the probability that individuals will interact with those with similar characteristics" (Greenberg & Mollick, 2017, p. 342), choice homophily is acting at the individual level. It explains how people choose to interact with similar others due to shared characteristics (Greenberg & Mollick, 2017). The concept of choice homophily has already been used to explain resource acquisition in different settings such as crowdlending (Riggins & Weber, 2017), venture capital investment (Hegde & Tumlinson, 2014), and crowdfunding (Greenberg & Mollick, 2017). However, it should be noted that homophily often was studied regarding only a single characteristic, such as *sex and gender*. For example, a first study in the context of crowdfunding demonstrated how female backers tend to support female entrepreneurs and even more so in industries where women are underrepresented (Greenberg & Mollick, 2017).

However, homophily is a multidimensional construct. It is not only related to sex and gender, but amongst others also to *race and ethnicity*, *age*, and *occupation* (McPherson et al., 2001). Therefore, it is necessary to include other, often neglected dimensions in the analysis to advance research on homophily, especially in online fundraising. Moreover, homophily occurs in the dyadic relationship between the investor and the entrepreneur in the online fundraising context. Especially in crowdfunding, however, research has so far focused primarily on the entrepreneur, neglecting the role

DOI: https://doi.org/10.5282/jums/v8i1pp237-269

of the investors. By studying how investors choose to support entrepreneurs, this paper aims to put the investors at the center of research.

Since homophily is a multidimensional construct in theory, it is of concern to adopt a multifaceted view in the online resource acquisition context to advance the understanding of the investors' decision-making. Moreover, it is simply not clear whether the findings for the characteristic sex are reproducible for other dimensions of homophily, such as race and ethnicity, age, occupation, or location. While one possible outcome of the analysis might be that investors exhibit homophily with regard to other dimensions, different mechanisms may just as well be responsible for the behavior of investors. In this context, the concept of intragroup competition (Deutsch, 1949; Goldman, Stockbauer, & McAuliffe, 1977), occurring due to perceived conflicting goal attainments among members of a group, could provide an alternative explanation for the behavior of investors. This scenario is particularly imaginable in the dimension occupation. Backers who have launched a crowdfunding campaign themselves may not fund projects in the same category as theirs since this may decrease the probability of success of their own campaign. Due to a lack of research to date, it is simply unclear to what extent homophily influences the investors' decision-making in online fundraising. Thus, the research question underlying this paper is as follows: *To what extent does homophily play a role in the choice of crowdfunding investors, and specifically, to what extent do different sociodemographic characteristics influence homophily?*

To answer the research question, I perform quantitative analyses on data gathered from Indiegogo, one of the largest crowdfunding platforms worldwide. For this purpose, I constructed a dataset of roughly 3.5 million backings made by more than one million backers on Indiegogo. Using a self-built web crawler, I collected various data points such as the name, location, profile image, and social media links of the backers and entrepreneurs. Thereby, my quantitative analyses focus on studying homophily effects in the five sociodemographic dimensions age, sex, race, occupation, and location. The different research hypotheses guiding my analyses can be summarized as follows: Investors are more likely to support those entrepreneurs to whom they are similar.

This study makes four contributions. First, I contribute to the broad literature on homophily. While scholars often studied homophily regarding single dimensions, for example, in a study demonstrating how women tend to support women in crowdfunding (Greenberg & Mollick, 2017), I extend the frame of analyses and individually consider five different sociodemographic dimensions of homophily. Thus, this study is one of the first to give a broader overview of homophily in different dimensions to do more justice to the complexity of this theoretical construct. In this context, I find that, as previously demonstrated, individuals tend to make homophilous choices regarding the sociodemographic dimensions age, sex, occupation, and location. In contrast, homophily in the more complex dimension race is not as clear-cut. Instead, I find significant differences depending on the race considered, where my findings suggest that primarily black and Indian individuals make homophilous choices.

Second, I contribute to the crowdfunding literature by moving the investors into the spotlight of crowdfunding research. Scholars have previously called to put the investors in the spotlight, as so far, research has mainly focused on the characteristics of the campaigns and the entrepreneurs as drivers of success, while the backers have been neglected in the crowdfunding context (Bretschneider & Leimeister, 2017; Kuppuswamy & Bayus, 2018). Analyzing the decision-making behavior of the investors also offers valuable practical insights for potential campaign creators. In this context, my findings suggest that black and Indian entrepreneurs are discriminated against by every but their race.

Third, most studies in crowdfunding obtained their data from the platform Kickstarter (exemplary see Greenberg and Mollick (2017)). My study empirically contributes by collecting the data on the backers' investment behavior from the crowdfunding platform Indiegogo.

Fourth, this study contributes by replicating the findings observed on other platforms for resource acquisition online on the crowdfunding platform Indiegogo. While, for example, a study on the lending-based microfinance platform Kiva showed that lenders tend to support fundraisers of the same sex and projects within an industry similar to the one the lenders are active in (Riggins & Weber, 2017), it is not clear whether these results are projectable to other platforms like Indiegogo. That is because scholars have argued that platforms differ, and thus "patterns observed on one platform cannot be assumed to generalize to other platforms" (Dushnitsky & Fitza, 2018, p. 1). My results, however, help to replicate and validate previous findings from other platforms for online fundraising and thus contribute by providing an additional perspective to the question of platform generalizability.

2. Theory and Hypotheses

The following chapter provides a review of the theories motivating this study. This chapter starts with a review of the general theory of homophily (see chapter 2.1), followed by an introduction to homophily in the context of crowdfunding (see chapter 2.2). Finally, this chapter ends with a theoretical explanation, including the hypothesis development, for each of the five sociodemographic dimensions of homophily examined in this paper (see chapter 2.3).

2.1. Homophily

The term *homophily* was initially coined by Lazarsfeld and Merton (1954) to describe the tendency of individuals to associate with similar others and to provide a clear distinction from the complementary term *heterophily*, which denotes a tendency for people who differ in certain respects to interact. The observation itself that people bond with others who are similar to them is rather old. Already Aristotle stated in his Rhetoric and Nichomachean Ethics that people "love those who are like themselves" (Aristotle, 1934, p. 1371), or Plato explained in Phaedrus that "similarity begets friendship" (Plato, 1968, p. 837).

Today, the phrase "birds of a feather flock together" (Lazarsfeld & Merton, 1954, p. 37) is commonly used to summarize the concept of homophily. In general, this theoretical construct describes a psychological process suggesting that similarity promotes connection, whereby bonds between similar people occur more frequently than between dissimilar people (McPherson et al., 2001). However, it is important to stress that homophily does not only occur in one dimension, such as race and ethnicity. Instead, it is a multidimensional construct, where race and ethnicity, age, sex and gender, religion, education, and occupation are among the main dimensions where similarity fosters interaction (McPherson et al., 2001).

The tendency of individuals to associate with others based on similar characteristics is one of the most thoroughly studied concepts in the social sciences (Greenberg & Mollick, 2017). In the past, homophily has been used to explain a plethora of phenomena in this field, such as personal contact (Wellman, 1996), friendship (Kandel, 1978; Lazarsfeld & Merton, 1954; Verbrugge, 1977), and marriage (Alba & Golden, 1986; Kalmijn, 1998). The overarching finding of all studies is that, for example, friends, colleagues, and individuals in romantic relationships are more similar to each other on various dimensions than randomly selected members of a population (Kossinets & Watts, 2009). In this context, early studies related to homophily and network formation were conducted among small social groups, where ties between group members could easily be observed by ethnographers, such as people meeting at a cafeteria, behavior among school children, or small urban neighborhoods (McPherson et al., 2001). Thereby, studies have demonstrated homophily by psychological traits such as aspirations, attitudes, and intelligence, suggesting that a community of values plays a vital role in the formation of friendships (Richardson, 1940). Other studies in the early phase were able to demonstrate significant homophily by demographic factors like sex, race and ethnicity, education, and age (exemplary see Bott (1928); Loomis (1946)).

With advances in research methods, scholars conducted large-scale studies of network formation and homophily beginning in the 1970s (McPherson et al., 2001). By taking advantage of more advanced research methods, studies on adolescent friendships were able to show that social status significantly influences friendships, as adults tend to choose friends with an equal social status (Verbrugge, 1977). Furthermore, studies on the formation and dissolution of adult friendships showed that adults choose friends with similar attitudes and that an incongruence of attitudes and behaviors can lead to friendship breakup (Kandel, 1978). In addition, studies that uncovered the mechanisms of the closest bond of marriage were able to show that weddings in the United States of America are more likely to happen between individuals with ethnically related backgrounds (Alba & Golden, 1986). Another study worth mentioning examined the influence of local ties, i.e., the neighborhood in which people live, on personal networks (Wellman, 1996). This study provided significant results showing the importance of local ties in personal networks, emphasizing the importance of the local dimension in interpersonal contact (Wellman, 1996). While the studies presented above, representing only a tiny portion of the vast literature on homophily, may differ in their contexts and methods, the primary finding that runs like a thread through all of the studies is that "people's personal networks are homogeneous with regard to many sociodemographic, behavioral, and intrapersonal characteristics" (McPherson et al., 2001, p. 415).

From those studies, two important aspects should be highlighted. First, most studies on homophily refer to the level of individual interaction, which describes the dyadic relationship between two individuals, such as the study on ethnic marriage (Alba & Golden, 1986). Besides the dyadic relationship between two individuals, the community level, which assumes that peer groups play an essential role in people's behavior, forms the other main tradition of research on homophily (McPherson et al., 2001). Second, the majority of the studies consider only one of the dimensions of homophily (Block & Grund, 2014), such as the influence of race and ethnicity on marriage (Alba & Golden, 1986). However, some scholars also emphasize the multidimensionality in network formation and social structure by measuring their development on several characteristics. In this regard, a study of social structure formation was one of the first to include multiple factors such as education, sex, and many more in the analysis (Blau, 1977).

Furthermore, two types of homophily—induced homophily and choice homophily—need to be distinguished. These different types of homophily describe the two possible mechanisms in which the similarity of related individuals may be grounded. First, induced homophily is caused by the group composition (Kossinets & Watts, 2009; McPherson & Smith-Lovin, 1987) and describes structural mechanisms defining the probability of interaction with similar others (Greenberg & Mollick, 2017). Thus, even in the case of random choices within the group, the social ties would create similarity between individuals (Greenberg & Mollick, 2017), suggesting that the group composition in the system dictates the possibility for association (McPherson & Smith-Lovin, 1987). On the other hand, choice homophily—the focus of this paper—operates at the individual level, is driven by shared characteristics, and explains how people choose to support and interact with similar others due to personal preferences (Greenberg & Mollick, 2017; Kossinets & Watts, 2009). Thus, in the case of pure choice homophily, the group's composition would not affect similarity between connected individuals, but the interpersonal interaction would be based solely on preferences for the dyadic similarity between the connected individuals (Greenberg & Mollick, 2017; McPherson & Smith-Lovin, 1987). Specifically, interaction is based on individual choices, with interpersonal similarity being the cause of association with someone similar to oneself (Greenberg & Mollick, 2015).

One possible reason for choice homophily presented in the literature is that being akin to someone—e.g., sharing the same cultural background, being the same age, or speaking the same language—can lead to trust and solidarity between individuals (Block & Grund, 2014). In short, choice homophily is "the individual-level propensity to choose similar others" (McPherson & Smith-Lovin, 1987, p. 371). However, it should be noted that historically it has been challenging to separate choice homophily from induced homophily (Greenberg & Mollick, 2015). That is because the different forms often do not occur in isolation but rather in some sort of interaction (Kossinets & Watts, 2009; McPherson & Smith-Lovin, 1987). Thereby, "groups are arenas for tie formation" (McPherson & Smith-Lovin, 1987, p. 373), where a group can be as small as a school class or large like platforms for online venture funding. Thus, opportunity structures for interactions are influenced by the size, composition, and structure of the group (McPherson, 1982, 1983; McPherson & Smith-Lovin, 1982, 1986), describing how induced homophily shapes choice homophily as structural barriers set the context on the options individuals can choose from (Greenberg & Mollick, 2017).

Over time, the concept of homophily has been applied not only in sociological research but also in other fields, such as business and economics. For example, a study examining the formation of inter-organizational alliances found that various rules of affiliation such as homophily, accumulative advantage, and follow-the-trend influence network evolution (Powell, White, Koput, & Owen-Smith, 2005). Those results are consistent with the findings of a study on structural homophily and firm partnering choices, which found that homophily can frequently be used to explain organizational alliances (Ahuja, Polidoro Jr, & Mitchell, 2009). Interestingly, a study focusing on interactions between employees within an organization found that employees exhibit significant homophily in communicating with other employees, especially in large groups that offer more choice of interaction partners than smaller groups (Kleinbaum, Stuart, & Tushman, 2013). This finding already points to the relevance of choice homophily, especially in large groups. Therefore, the concept of choice homophily is particularly relevant in the context of online venture funding, especially crowdfunding, as it involves large groups, and studies in this context aim to explain how investors choose to fund specific projects and entrepreneurs. For example, in this setting, a study examining the influence of social proximity on business partnerships found that venture capitalists tend to choose start-ups with co-ethnic executives for their investment, describing an ethnic proximity between the venture capital firm and the start-up (Hegde & Tumlinson, 2014). Similarly, an initial study of homophily in the context of crowdfunding—which will be mentioned only briefly here and discussed in more detail in the following chapter—found that females tend to support female entrepreneurs (Greenberg & Mollick, 2017). The above studies show that the concept of choice homophily has gained interest in the business and economics contexts, where choice homophily has been used to explain various settings of entrepreneurial funding. Thereby, recent interest has also turned to crowdfunding markets, as they provide a near-optimal framework to focus almost solely on choice homophily (Greenberg & Mollick, 2017).

2.2. Homophily in Crowdfunding

As explained in chapter 2.1, homophily has already been applied in various fields of literature, ranging from sociology to business and economics. Focusing almost purely on choice homophily, an initial study in the context of crowdfunding demonstrated that women tend to support other women, and even more so in industries where women are typically underrepresented (Greenberg & Mollick, 2017). In this context, Greenberg and Mollick (2017) further divide choice homophily into two more detailed mechanisms in order to differentiate between homophily caused by mere dyadic similarity, i.e., interpersonal choice homophily, and homophily caused by "perceptions of shared structural barriers stemming from a common group-level social identity and an underlying desire to help overcome them" (p. 342), i.e., activist choice homophily.

The goal of the above study was to choose a context in which homophily has historically been viewed as a cause of inequality, namely the funding of female entrepreneurs (Greenberg & Mollick, 2017). Thereby, past studies pointed to the fact that women receive less venture capital than men. For example, women-led firms receive only a small percentage of venture capital investments (Greene, Brush, Hart, & Saparito, 2001; Harrison & Mason, 2007). A possible reason for this gender gap is the inequality in the traditional funding network itself, where less than 15 percent of venture capitalists are female (Harrison & Mason, 2007; Stuart & Sorenson, 2007), having scholars suggest that homophily is one of the reasons for the gender imbalance in venture funding (Becker-Blease & Sohl, 2007; Stuart & Sorenson, 2007). However, connecting to the overarching theory of homophily, it is worth noting that an inequality exists not only in the funding of female entrepreneurs but also in the broader context of resource acquisition concerning other sociodemographic characteristics such as race and ethnicity. In this context, studies have shown that inequality also exists in this regard, as African American men are less likely than white entrepreneurs to obtain funding (Jenq, Pan, & Theseira, 2015; Younkin & Kuppuswamy, 2018). This observation may be due, at least to some degree, because most early-stage investors are white men (Sohl, 2015). To study such biases and to focus merely on the choices made by investors, that is, on the concept of choice homophily, crowdfunding provides a more appropriate context than traditional means of funding such as venture capital firms (Greenberg & Mollick, 2017). Thus, the crowdfunding context offers a solution to the long-standing challenge of separating induced and choice homophily (Greenberg & Mollick, 2015, 2017).

The term crowdfunding was first coined in the late 2000s and "describes a new institutional form which utilizes digital platforms to originate and aggregate funding" (Dushnitsky & Fitza, 2018, p. 1). While early-stage funding and resource acquisition have always been crucial for commercializing innovative ideas (Cosh, Cumming, & Hughes, 2009; Kortum & Lerner, 2000), deciding which new ideas to commercialize has long been tremendously undemocratic, with small groups of experts, typically white male venture capitalists, deciding which new idea to support (Mollick & Robb, 2016). The lack of diversity in traditional venture funding and the resulting barriers in access to financial resources (Mollick & Robb, 2016) have led to an increasing number of entrepreneurs using a relatively new form of financing that makes use of large online communities, known as *crowdfunding* (Kuppuswamy & Bayus, 2018; Mollick, 2014). In a prevalent definition of the term, crowdfunding "refers to the efforts by entrepreneurial individuals and groups (...) to fund their ventures by drawing on relatively small contributions from a relatively large number of individuals using the internet, without standard financial intermediaries" (Mollick, 2014, p. 2). In this context, the relatively large number of individuals is also commonly referred to as *the crowd* (Belleflamme, Lambert, & Schwienbacher, 2014). Appealing to a large community of investors from the general public, crowdfunding "involves an open call, mostly through the Internet, for the provision of financial resources either in the form of donation or in exchange for the product or some form of reward to support initiatives" (Belleflamme et al., 2014, p. 4). Ultimately, the goal of crowdfunding platforms is to connect entrepreneurs with investors who are willing to support new ideas (Bretschneider & Leimeister, 2017). Thereby, the goal is to democratize the funding process by serving as a vehicle for greater participation by investors and entrepreneurs who are typically underrepresented in these markets, such as women and all sorts of minorities (Mollick & Robb, 2016). While crowdfunding can currently still be considered a niche phenomenon in total funding volume, it is gaining importance in many countries (Belleflamme, Omrani, & Peitz, 2015). The crowdfunding market has grown from US$880 million in 2010 to US$34.4 billion in 2015 (Massolution, 2015) and is expected to grow to US$96 billion by 2025 (World Bank, 2013). Although these figures are comparatively small compared to the trillions of dollars invested overall, they show that crowdfunding is a rapidly growing market (Belleflamme et al., 2015).

As can be inferred from the brief description above, the crowdfunding context is particularly well suited to focus on the concept of choice homophily. By reducing social constraints and search costs, crowdfunding allows campaign creators to access potential backers with similar sociodemographic characteristics and allows individual investors to choose to fund ideas they support (Greenberg & Mollick, 2017). Thus, choosing the crowdfunding context can lower induced homophily in two main ways and allows to focus almost entirely on choice homophily. First, it removes the traditional gatekeepers of venture funding, such as venture capitalists, and at the same time increases the set of potential investors by offering everyone the opportunity to fund a campaign (Greenberg & Mollick, 2017). As shown above, traditional gatekeepers to start-up funding lack diversity with respect to various sociodemographic characteristics, such as sex or race. Thus, since the lack of diversity limits the possibility for connections between investors and entrepreneurs, traditional funding includes high levels of induced homophily, disfavoring minorities that have historically been disadvantaged in the venture funding process. However, crowdfunding potentially democratizes resource acquisition by removing traditional gatekeepers and providing investment opportunities to a possibly unlimited community of investors from the general public. It increases diversity among investors with respect do various sociodemographic characteristics, ultimately reducing induced homophily.

Second, using digital platforms as a means of funding makes it possible to reach a larger group of like-minded individuals than is typically the case with socially or geographically constrained searches (Greenberg & Mollick, 2017). In line with McPherson and Smith-Lovin (1987), arguing that given a certain amount of diversity, larger groups allow for more homophilic ties, crowdfunding, with its inherently large and diversified community, presents the perfect setting to focus on individual choices. In this setting with a large set of possible choices, it is thus possible to properly study choice homophily because if individuals "make homophilous choices, they produce similar pairings even in cases of high group diversity" (McPherson & Smith-Lovin, 1987, p. 372).

2.3. Sociodemographic Dimensions of Homophily

The following chapter aims to theoretically motivate and explain the five sociodemographic dimensions of homophily studied in this paper. I provide a theoretical explanation regarding the backers' possible behavior with respect to the dyadic relationship with the entrepreneur for each dimension analyzed—i.e., *age*, *sex*, *race*, *occupation*, and *location*—including the respective hypotheses guiding the analyses. In line with the goal of this paper, which is to examine how investors choose to support entrepreneurs in online resource acquisition, the following chapter focuses on choice homophily. Thus, this chapter provides theoretical explanations for situations where choice is warranted, describing why investors might show a preference to invest in entrepreneurs to whom they are similar.

2.3.1. Age

Apart from generation-linking ties such as relationships with children or parents, age homophily has been proven to be quite strong across general social relationships (Smith, McPherson, & Smith-Lovin, 2014). However, typically a large part of age homophilic networks is induced by institutional settings and life-course patterns such as workplaces, schools, or sports clubs (Kalmijn & Vermunt, 2005). These findings are consistent with a study by Feld (1982), who argues that the environments in which people engage tend to be homogeneous, indicating that the group's composition implies the extent of age homophily in a network.

Apart from age homophilic networks induced by structural and institutional environments, scholars have also presented mechanisms that explain the emergence of age homophilic networks due to individual choice. For example, a study in the social network context was able to transfer the findings of age homophily from the offline world to the online world, specifically to the social network Myspace (Thelwall, 2009). This finding is interesting because although the internet is known for increasing the diversity of opinions, information, and sociodemographic characteristics of the pool of an individual's potential contacts, the study shows that people still choose to connect with similar others (Thelwall, 2009). Thus, the results of this study suggest that age homophily is not only induced but also a result of choice.

In addition, scholars have argued that even minor age differences can lead to significant differences in individuals' interests (Smith et al., 2014), suggesting that individuals choose to interact with other individuals who are similar in age based on shared interests. This finding is consistent with studies on age differences and changes in activities, demonstrating a shift in activities as people age due to changes in individual preferences, abilities, and constraints (Verbrugge, Gruber-Baldini, & Fozard, 1996). This idea also seems plausible in the context of online resource acquisition. It is conceivable that individuals of a similar age share similar interests. Therefore, an investor might be particularly attracted to the product or service of an entrepreneur of similar age because they both share the same interests, possibly reflected in the product or service.

In addition, the theories of social identity (Tajfel, 1974) and self-categorization (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987) provide a possible explanation for the investors' age-homophilic behavior. The process of classifying individuals into similar others (in-group members) and dissimilar others (out-group members) is a basic psychological process rooted in the desire to increase self-esteem. Thereby individuals tend to exhibit a bias in favor of members of their group (Avery, McKay, & Wilson, 2007). This categorization of others and oneself and associating values with the various social categories (Tajfel, 1974; Turner & Oakes, 1986) is related to the goal of individuals to maintain persistent identities (Steele, 1988), which explains why individuals tend to positively value similar individuals over dissimilar individuals (Goldberg, 2005). Please note that the above paragraph considers the theories therein in light of the sociodemographic dimension age. However, these theories can, in principle, explain the behavior in terms of a variety of different demographic attributes defining individuals, as according to those theories, individuals classify others and themselves based on various dimensions such as age, gender, or race (Avery et al., 2007). Therefore, I will also refer to what is presented here in the following chapters to explain possible outcomes in other dimensions.

Consistent with this line of reasoning, the framework of relational demography (Tsui & Gutek, 1999; Tsui & O'reilly, 1989) argues that individuals act more favorably in environments with a higher number of in-group members in terms of demographic characteristics like age. Thus, similarity in demographic characteristics increases perceived similarity in experiences and values, thus promoting coherence and identification among in-group members (Mehra, Kilduff, & Brass, 1998). Additionally, the similarity-attraction paradigm (Byrne, 1971), closely related to the social identity theory, should also briefly be mentioned here. This paradigm argues that similar individuals are attracted and experience positive outcomes in interaction and association (Goldberg, 2005).

Similarly, it seems plausible that investors in online resource acquisition tend to support entrepreneurs of similar age, both because of shared interests, possibly leading to a high commonality of interest in the funded project, and because of self-categorization and social identity. Therefore, I propose:

> *Hypothesis 1 (H1): Contributions to entrepreneurs with a similar age as the investor occur more frequently than we would expect if investors chose projects for contribution at random (baseline).*

2.3.2. Sex and Gender

As already mentioned, an initial study on homophily in the context of crowdfunding was able to show that women choose to support each other (Greenberg & Mollick, 2017). In this context, a vital driver of the favored behavior regarding similar individuals is the basic principle of choice homophily, which states that individuals, on average, have a positive affection for common characteristics (Huston & Levinger, 1978; Ingram & Morris, 2007; Lazarsfeld & Merton, 1954; Verbrugge, 1977). Scholars propose different drivers of choice homophily in this context, where these drivers include, among others, trust and ease of communication (Kossinets & Watts, 2009; Wimmer & Lewis, 2010). However, please note that the principle of choice homophily—attraction due to shared characteristics—can be used to explain the behavior in terms of any sociodemographic characteristic shared by two individuals (Wimmer & Lewis, 2010). Thus, the principle of choice homophily represents a central pillar for explaining the possible behavior of investors, which I will refer to throughout chapter 2.3.

The preference for shared characteristics, commonly used to explain the emergence of same-sex ties, is consistent with what is presented for the theories of social identity and self-categorization (see chapter 2.3.1). Those theories describe how individuals classify themselves and others and prefer ties to in-group members in part to increase self-esteem (Avery et al., 2007). In the context of increasing self-esteem, another argument seems plausible to describe the tendency of individuals to support others of the same gender in the process of resource acquisition, which is that individuals "seek information affirming identification with their in-groups" (Avery et al., 2007, p. 1543). This argument suggests that individuals do not blindly look for similarity, such as in terms of gender, as a basis for engagement but instead look for identity-affirming similarity (Avery et al., 2007). In short, this means that the bias only occurs when the members of the in-group that are potentially interacted with help maintain and create a positive impression of that group (Avery et al., 2007). This argumentation is particularly applicable to resource acquisition online because, in this context, investors choose to support entrepreneurs, where the entrepreneur is generally positively portrayed as an individual who aims to create wealth and foster innovation (Carland, Hoy, & Carland, 1988).

In addition, the same-gender bias effect (Mobley, 1982) could provide a possible theoretical explanation for the behavior of investors in choosing whom to support or not support in online venture funding. This principle has shown that, for example, in an organizational setting, raters rate members of their gender subgroup higher than members of other subgroups (Mobley, 1982). Similarly, it is conceivable that not only in organizational settings but also in online resource acquisition, raters (i.e., the investors) tend to value members of their gender subgroup higher than members of other subgroups and therefore show a tendency to support entrepreneurs with whom they share the same gender.

In line with the results expected from the above argumentation, several studies in the context of crowdfunding were also able to demonstrate that investors show a tendency to fund entrepreneurs of the same gender (Greenberg & Mollick, 2015, 2017; Groza, Groza, & Barral, 2020; Riggins & Weber, 2017; Vismara, Benaroio, & Carne, 2017). Thus, in line with the theories presented above as well as findings from previous studies, I propose:

> *Hypothesis 2 (H2): Contributions to entrepreneurs of the same sex as the investor occur more frequently than we would expect if investors chose projects for contribution at random (baseline).*

2.3.3. Race and Ethnicity

Race and ethnicity undeniably present the most significant divide in social networks (McPherson et al., 2001). Thereby, research has demonstrated significant racial homophily in a variety of relationships ranging from school friendships (Shrum, Cheek Jr, & MacD, 1988) and work relations (Ibarra, 1995; Lincoln & Miller, 1979) to the close bond of marriage (Kalmijn, 1998). It should also be noted that race and ethnicity are among the sociodemographic characteristics that lead to the highest levels of choice homophily (McPherson et al., 2001). For example, a study of interracial friendships in secondary schools found that cross-racial friendships were only one-sixth as likely as choosing a same-race friend (Hallinan & Williams, 1989).

Particularly relevant in the context of resource acquisition online are findings that show that anxiety among members of the racial majority increases when the proportion of racial minorities in the group increases (Stephan & Stephan, 1985). This change in the group composition was shown to cause members of the majority group to experience a threat to their superior status (Abrams & Hogg, 2006; Tajfel, 1974). While the white population has historically been the majority group in the startup funding process, online resource acquisition platforms—such as crowdfunding—democratize access to venture funding (see chapter 2.2) and thus increase the proportion of racial minorities. As a result, this change could lead to a perceived threat to the formerly mighty group of white individuals, explaining why white investors could continue to show a tendency to support white entrepreneurs. However, these findings only illustrate why homophily might occur among majority group members in venture funding, i.e., the white population historically. This does not provide a possible reason for homophily in the dimension race and ethnicity in general. In this light, it is also important to mention a study on racial discrimination in crowdfunding, showing that black entrepreneurs are discriminated against (Younkin & Kuppuswamy, 2018), which also only distinguishes between the white population (majority group) and the black population (minority group). However, a broader analysis is required since there are various races besides only white or black individuals. Therefore, the following section explains why it is conceivable that homophily occurs in general in the dimension race.

Just as for the two sociodemographic characteristics presented in the previous chapters, the theories of self-categorization and social identity (see chapter 2.3.1) could explain the decision-making behavior in online resource acquisition with respect to the dimension race and ethnicity. As explained earlier, individuals classify themselves and others into in-group and out-group members. Furthermore, due to the desire to increase their self-esteem, they tend to react negatively to threats to their social identity, such as discrimination (Avery et al., 2007). This concept seems plausible to also explain why homophily effects might occur with respect to race and ethnicity. According to this concept, investors tend to support projects from entrepreneurs of the same race to prevent discrimination against their race and thus reduce threats to their social identity.

The principle of choice homophily, where individuals choose to interact with similar others solely based on dyadic similarity (Greenberg & Mollick, 2017; Kossinets & Watts, 2009; McPherson & Smith-Lovin, 1987), furthermore provides a general explanation for why investors could support entrepreneurs of the same race in crowdfunding. The idea of choice homophily is further supported by interesting results of several studies showing that people generally recognize faces of their race better than faces of other races (Brigham & Barkowitz, 1978; Malpass & Kravitz, 1969). This same-race bias seems especially relevant in crowdfunding. In this context, it is conceivable that an investor easily recognizes an entrepreneur of the same race based on the profile picture and might therefore, in line with the concept of choice homophily, choose to support the project of an entrepreneur with the same race over the project of an entrepreneur with a different race.

Furthermore, a first study in the crowdfunding context was able to demonstrate racial homophily as investors tend to fund projects from entrepreneurs in the same racial group (Dahlin, Rhue, & Clark, 2019). Thus, according to the theories presented above as well as in line with previous findings, I propose:

> *Hypothesis 3 (H3): Contributions to entrepreneurs of the same race as the investor occur more frequently than we would expect if investors chose projects for contribution at random (baseline).*

2.3.4. Occupation

Contrary to what has been presented for the previous dimensions, the literature provides arguments opposing the idea of homophily in the dimension occupation, suggesting that investors might notably not support entrepreneurs in the same industry as they are active in. In this context, it is worth mentioning the concept of intragroup competition (Deutsch, 1949; Goldman et al., 1977), which describes how competition arises within a group caused by perceived conflicting goals among the members of the same group. While the concept has previously been used, for example, to explain how members of an organizational department compete for the largest share of a fixed annual budget (Goldman et al., 1977), it also seems appropriate for explaining the behavior of those investors in online resource acquisition who are entrepreneurs themselves. In this context, it seems conceivable that individuals who have themselves launched a campaign on the platform might not want to invest in campaigns within their industry (intragroup), as they might assume that this would reduce the probability of success of their campaign. This seems plausible as investors are generally influenced by the amount of funding that has already been pledged to a project (Kuppuswamy & Bayus, 2018). Thus, the entrepreneur-investor might assume that funding the project of another (potentially competing) entrepreneur in the same industry could make other investors prefer the competitor's campaign, as they could be influenced by the potentially higher sum already pledged to the competitor's project. Ultimately, this could lead entrepreneur-investors not to support projects of other entrepreneurs operating in the same industry so as not to promote competition with respect to their own projects.

However, this is countered by the extensive literature on homophily. According to McPherson et al. (2001), significant homophily has been found in the past for the dimension occupation. For example, in a study in the context of adult friendship choice, similarity in profession was almost equally important and influenced the formation of a tie almost as much as sharing the same gender (Verbrugge, 1977). The sociological finding that individuals choose friends with a similar occupation more often than chance would suggest is explained in part by the principle of status similarity. In this context, a similar social position, for example, in terms of occupation, is valuable because this indicates similar experiences and viewpoints (Verbrugge, 1977). While the previous study examined choices based on similarities in occupation in the friendship context, the logic also seems conceivable in online resource acquisition, as scholars argue that shared experiences and viewpoints also exist in this context. Thereby, especially in crowdfunding, investors often act as campaign creators themselves and can therefore relate to the challenges of their peers, which argues for comradery among this group of people (Colombo, Franzoni, & Rossi–Lamastra, 2015; Groza et al., 2020). In this light, the concept of social capital, generally defined as "the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit" (Nahapiet & Ghoshal, 1998, p. 243), is used to explain why investors who are entrepreneurs themselves in general tend to support other entrepreneurs in online venture funding. Social capital includes both the network and the resources mobilized through that network (Bourdieu, 1986) and describes, for example, how shared values and understandings enable individuals to trust each other and thus to cooperate (Nahapiet & Ghoshal, 1998). In this regard, scholars argue for the importance of internal social capital—that is, the contacts established within the crowdfunding platform (Colombo et al., 2015). They argue for a perceived obligation of the entrepreneur to give back to those who helped fund the campaign (specific reciprocity) and a perceived duty to support other campaigns because entrepreneurs are grateful to have received funding in the past (generalized reciprocity). Consistent with this, Colombo et al. (2015) show that investors who are entrepreneurs themselves generally tend to support other entrepreneurs in crowdfunding because they share common challenges, experiences, and a sense of mutual identification.

However, this sense of mutual identification and shared experiences is not limited to the high-level entrepreneur-to-entrepreneur relationship. It is also expected to occur one level deeper in the industry-to-industry relationship, implying that not only do entrepreneurs support each other but that positive homophily effects are also expected between entrepreneurs in the same industry, such as the technology sector. Consistent with the shared values and experiences presented in the social capital literature, this argumentation is further supported by a study in the venture capital context, which shows that industry specialization of the venture capitalists leads to better expertise and thus better investments (Gompers, Kovner, & Lerner, 2009). Since investors do not want to invest in projects that fail, and because of similar experiences and viewpoints, it seems plausible that also in the context of resource acquisition online, investors will invest in campaigns from entrepreneurs that are active in the same industry as themselves.

Consistent with the above argumentation, an initial study in the context of online resource acquisition, specifically crowdlending, was able to show that lenders tend to support projects in the same industry as they are active in (Riggins & Weber, 2017). Thus, following the theories presented above as well as the principle of choice homophily—association due to shared characteristics—I propose:

> *Hypothesis 4 (H4): Contributions to entrepreneurs with the same occupation as the investor occur more frequently than contributions to entrepreneurs with a different occupation than the*

investor.

2.3.5. Location

While all four characteristics presented in chapters 2.3.1 to 2.3.4 are attributed in the literature to the sociodemographic dimensions of homophily that stratify society (McPherson et al., 2001), location is often referred to less as a dimension and more as a source of homophily in the mainstream literature (McPherson et al., 2001). The main reason why geography is considered a source of homophily in the traditional sociological literature is that individuals are simply more likely to interact with individuals who are spatially closer than with individuals who are further away (McPherson et al., 2001). This finding at the same time indicates a high degree of induced homophily with respect to this demographic attribute. In line with this, one study argues, for example, that personal networks are local, suggesting that local contacts are an essential source of routine interactions (Wellman, 1996). This argumentation provides a plausible explanation for traditional sociological research, primarily concerned with forming ties in face-to-face scenarios, with residence inducing a particular frame of possible contacts. In the context of online resource acquisition, however, the situation is fundamentally different, as it is suitable to break the limitation of space. In this case, the geographic location does not seem to be a source of homophily in the traditional sense, as the use of digital platforms removes the direct influence of location on the likelihood of two people interacting. Thus, the location of an individual does not induce the set of possible connections in the online funding process and therefore allows to solely focus on choice homophily with regard to the geographic location. Therefore, it is relevant to investigate whether the location impacts the investors' choice regarding whom to fund. Thus, location is not considered a source but rather a dimension of homophily in my study.

It seems plausible that investors might show homophilous behavior towards entrepreneurs whom they are geographically close to, based on the concept of local bias, which states that investors tend to invest in geographically close companies (Coval & Moskowitz, 1999). Researchers have identified several reasons why investors choose to invest in companies that are close to them. These reasons include, for example, more accessible information about nearby companies, with investors obtaining information from local media or even from local relationships with people from nearby companies (Coval & Moskowitz, 1999, 2001). Other reasons for geographically proximate investments mentioned by scholars are more psychological, such as a desire to fund the local community (Coval & Moskowitz, 1999) or investors simply feeling more comfortable about local firms (Coval & Moskowitz, 1999; Huberman, 2001).

In the past, several studies have shown that the local bias is present in different organizational contexts. For example, studies examining the influence of distance on portfolio choice showed that fund managers in the United States prefer stocks of companies located nearby (Coval & Moskowitz, 1999, 2001). Moreover, consistent with previous findings, an entrepreneurial finance study also showed that venture capital firms exhibit a significant local bias (Cumming & Dai, 2010). Likewise, it seems plausible that geographic proximity is relevant not only in the corporate context but also for private investors, such as in crowdfunding.

In line with this argumentation, the home bias phenomenon, being closely related to local bias, provides further theoretical explanations why investors in online resource acquisition could tend to invest in entrepreneurs whom they are geographically close to. The home bias describes a phenomenon where transactions are conducted among geographically close parties, for example in the same country or state (Lin & Viswanathan, 2016). While home bias has frequently been studied in offline contexts, a study in the context of online auction sites, such as eBay, was able to show that transactions are still more likely to happen between sellers and buyers who are geographically close (Hortaçsu, Martínez-Jerez, & Douglas, 2009). Even though those marketplaces are online, the authors argue that the geographic location is still influential due to location-specific goods like event tickets, shipping charges, or the possibility of direct contract enforcement in case of breach (Hortaçsu et al., 2009).

The results from the studies mentioned above have also been proven in online venture funding, where studies have shown that also in crowdfunding, backers tend to invest locally (Giudici, Guerini, & Rossi-Lamastra, 2018; Guo, Guo, Wang, Wang, & Wu, 2018; Lin & Viswanathan, 2016). Therefore, in line with the above argumentation as well as previous findings, I propose:

> *Hypothesis 5 (H5): Contributions to entrepreneurs with a similar location as the investor occur more frequently than we would expect if investors chose projects for contribution at random (baseline).*

3. Data and Methods

The following chapter provides a description of the data and methods used throughout my analyses. The chapter starts with a brief introduction to the setting of this study (see chapter 3.1), followed by an introduction to the sample, including an explanation of the data collection process (see chapter 3.2). Subsequently, the various variables used in this study are presented (see chapter 3.3). Finally, this chapter ends with the sample's descriptive statistics (see chapter 3.4) and an outline of the statistical methods used in my analyses (see chapter 3.5).

3.1. Setting

To analyze homophily, I selected the crowdfunding platform Indiegogo as the setting of this study. Indiegogo is the second-largest crowdfunding platform after Kickstarter. It belongs to the group of reward-based crowdfunding platforms, where investors receive a reward in return for backing a project with a certain amount of money (Mollick, 2014).

Examples of these rewards—also referred to as *perks* on Indiegogo—include early access to products, the possibility to meet the project creators, or being credited in the supported project (Mollick, 2014).

Founded in 2008 with the mission to empower people to support ideas that matter to them and collectively make those ideas come to life, Indiegogo, with its community of over nine million backers from about 235 countries, has already helped to realize over 800,000 innovative projects (Indiegogo, 2021a). Today, roughly 19,000 campaigns are launched on Indiegogo every month, and about ten million people from all over the world visit the platform during the same time. It is interesting to note that on Indiegogo, women launch 47% of the campaigns exceeding their funding target, which points towards a democratization of the funding process. In line with what has already been presented in chapter 2.2 for crowdfunding platforms in general, Indiegogo as one of those platforms offers the perfect setting to study biases in the choice and decision-making of investors—i.e., choice homophily, the focus of my study. That is because thousands of projects in various categories are offered on Indiegogo as potential investment opportunities for backers from all over the world, where those backers in principle could choose from every possible project, thus lessening induced homophily to a minimum.

In principle, the idea of Indiegogo can be explained as follows. Indiegogo offers a comprehensive platform for entrepreneurs to make their projects available to a large set of potential investors. In return, Indiegogo charges a five percent platform fee on the actual funds raised for the campaign (Indiegogo, 2021d). Entrepreneurs can choose between a fixed funding model (all-or-nothing) or a flexible funding model (keep-what-you-raise). The fixed funding model is used mainly for projects where a strict threshold must be met to realize the project (such as typically in manufacturing). All contributions are returned to the investors if the project does not meet its goal. On the other hand, the flexible funding model, where the entrepreneurs receive all contributions even if the campaign goal was not met, is better suited for projects where no strict minimum financing is required to realize the project. Furthermore, in principle, a campaign on Indiegogo can run for up to 60 days. However, Indiegogo advises a campaign duration of about 40 days, amongst others, for reasons of momentum, engagement, and urgency (Indiegogo, 2021b).

Furthermore, each campaign is assigned to a specific category on Indiegogo. In principle, projects are classified according to the three top-level categories *Tech & Innovation*, *Creative Works*, and *Community Projects* (Indiegogo, 2021c). Each of the three top-level categories comprises a group of more detailed categories. For example, the top-level category *Tech & Innovation* includes categories such as Phones & Accessories, Transportation, Home, or Productivity. The top-level category *Creative Works* contains categories such as Art, Music, or Video Games. Appendix 1 includes a detailed overview of all Indiegogo categories and their allocation to the corresponding top-level categories. Please note that in the following, I will use the term 'top-level categories' when talking about Tech & Innovation, Creative Works, and Community Projects, and I will use the term 'categories' when talking about the more detailed categories.

The individuals who are funding projects on the platform are commonly referred to as *backers* on Indiegogo (Indiegogo, 2021a). When describing the results from my quantitative study, I will refer to this group of people as *backers* or *investors*. The individuals requesting funds are commonly referred to as *entrepreneurs* on Indiegogo (Indiegogo, 2021a). When describing the results from my analyses, I will refer to this group of people as *entrepreneurs* or *founders*.

3.2. Sample

I created a unique portfolio of investments made by backers on the crowdfunding platform Indiegogo. The overall dataset consists of 3,509,077 investments into projects launched between January 2008 and February 2021. Those investments were made by 1,202,233 backers who invested in 171,116 different projects initiated by 147,780 entrepreneurs.

Each observation in my dataset—representing one investment made by one backer—consists of three sections in terms of content. The first section contains various personal information about the backer, such as the name, location, profile picture, personal description, and links to connected social networks. The second section contains various information about the campaign funded, including the category, the amount of funding, the duration of the project, and information about the number of investors, comments, and updates of the project. Finally, the third section contains—similar as for the backer—personal information about the entrepreneur who launched the funded campaign. Therefore, each observation in my dataset represents one dyad, that is, one backer-to-entrepreneur relationship.

I obtained the complete contribution history for all backers in my dataset between January and March 2021. The data collection process can be described as follows. From an initial set of projects launched on Indiegogo, I extracted the unique identifiers (IDs) of all backers who funded those projects using a self-built Python script. This list of roughly 1.2 million different backers represents the actual start of my data collection process. For this purpose, I built a web crawler to collect the portfolios of the individual backers, including all information needed to study the five dimensions of homophily introduced throughout chapter 2.3. To develop my web crawler, I used the open-source framework Scrapy[1]. The process of constructing the portfolio of the backers using my self-built web crawler can be described as follows. First, using the backer IDs as input, I extracted all available personal data about the backers from their Indiegogo profile page (e.g., their name and location, as described above). Second, since Indiegogo profile pages contain not only personal data about the individual but also the history of their

[1] Scrapy is a high-level web crawling framework for Python, used for fast extraction of data from websites (Scrapy, 2021).

activity on the platform, I extracted all contributions made by the backers.

Furthermore, from crawling the funding activity of the backers, I obtained both the IDs of the projects funded and the IDs of the entrepreneurs who launched the corresponding projects. Using this additional information, I collected various data about the projects funded (this data includes, as described above, information about the project category, the funding duration, and the amount of funding). As the last part of my data collection process, I used the same web crawler I already used for the backers for all entrepreneurs who launched the projects to extract all available personal data about the entrepreneurs. By merging the various parts, I finally got to my overall dataset containing about 3.5 million investments made on Indiegogo.

Please note that due to differences in data availability, I created subsamples for my analyses. Those subsamples (i.e., subsets of the overall dataset) were created depending on the availability of the data required for the analysis in the respective dimension. For example, for the analysis in the dimension location, only those observations were considered for which both the investor and the entrepreneur indicated a location on their Indiegogo profile page. The same holds for the other dimensions. For example, only those observations were included in the sample used to analyze racial homophily, where a race was identified for both the backer and the founder. A particular case lies in the creation of the subsample in the dimension occupation. Since the relationship between investors who are entrepreneurs on Indiegogo themselves and the entrepreneurs they fund is examined in this case, the corresponding sample was limited to all those observations where the investor has also launched at least one own campaign. Table 1 shows the number of dyads (# Observations), the number of unique backers (# Backers), the number of unique projects (# Projects), as well as the number of unique founders (# Founders) in the overall dataset as well as for each of the subsamples used in my analyses.

Furthermore, it should be mentioned that while the samples from table 1 contain all investments of all backers (and thus potentially several observations per backer if more than one investment was made by an investor), the various analyses were performed considering one (aggregated) observation per investor. This normalization at the backer level was performed after the calculations required to construct the dependent variables, where the activities of an individual backer were aggregated into special measures per dimension (see chapter 3.3.1). Thus, the sample sizes of my different models are indicated by the number in the column "# Backers" in table 1. However, please note that the final sample sizes were reduced again due to the data availability of the various control variables used in the analyses. Therefore, the final sample sizes are indicated in the figures of the respective models throughout chapter 4.

3.3. Variables

This chapter describes the variables used in my analyses. The chapter starts with a description of the dependent variables (see chapter 3.3.1), followed by the independent variables (see chapter 3.3.2). Afterward, the control variables used in the analyses are presented (see chapter 3.3.3).

3.3.1. Dependent Variables

My basic approach to analyzing homophily in crowdfunding can be summarized as follows. To test for the presence of homophily, I test for deviations of an investor's actual backings from the expected backings—describing what portfolio composition would be expected if the investor chooses ties randomly, i.e., the baseline. This approach is closely related to the example of a person rolling a dice several times. If the die is not loaded (comparable to random choice in my setting), the average of the actual numbers rolled (actual backings) should equal the expected value of the die roll (baseline). However, suppose the average of all the actual numbers rolled significantly deviates from the expected value. In that case, we could suspect a loaded die (i.e., biased choice in my context, which, depending on the interaction with the respective independent variable, indicates homophily).

For further clarification, the approach is now illustrated using a simplified example of the dimension age (compare H1). Let us assume that among all investments made in the crowdfunding universe, the average age of the entrepreneurs funded is 30 years. This average age represents the simplified baseline. Let us now consider the portfolio of one fictitious investor. We expect that the average age of all entrepreneurs funded by this investor is also around 30 years if ties were selected randomly (i.e., if homophily would not be present). Now consider the actual portfolio of this fictitious investor who has backed three projects in the past. One entrepreneur funded by the investor is 19 years old, the second is 20 years old, and the third is 21 years old. Thus, by calculating the mean age of the entrepreneurs in the investor's actual backings, we find that the actual average age of the entrepreneurs funded by the investor is only 20 years. Comparing both numbers indicate a deviation of the backer's actual portfolio from the expected portfolio, whereby the investor, on average, supports younger entrepreneurs than chance would suggest. Depending on whether the backer is younger or older himself, we could ultimately argue for or against homophily.

Therefore, in general, two measures are necessary to construct the dependent variables, which denote the deviation of a backer's actual investments from the baseline. One is the actual backings of an investor, and the other is the expected backings, that is, the baseline. The baseline denotes the expected portfolio of an investor depending on the categories of the projects funded by an investor. The actual backings, on the other hand, represent the actual portfolio of a backer. Thereby, a portfolio contains all investments that an investor has made at the time of the data collection. However, there are slight differences in the type and measurement of the dependent variables, depending on the particular hypothesis of my study considered. The following section thus begins with the description of the dependent variable for the analysis in the dimension age (H1), followed by the dependent variables for the dimensions sex (H2), as well as race (H3).

Table 1: Overview of the overall dataset and different subsamples

Source: Own illustration

Sample	# Observations	# Backers	# Projects	# Founders
Overall	3,509,077	1,202,233	171,116	147,780
Sex (DeepFace)	8,337	3,563	4,315	3,845
Sex (Genderize.io)	2,423,919	945,651	151,364	130,949
Age	8,337	3,563	4,315	3,845
Race	8,337	3,563	4,315	3,845
Location	401,967	109,241	51,771	44,242
Occupation	104,590	27,920	50,250	44,701

Subsequently, the dependent variables for the dimension location (H5) are presented. Please note that the construction of the dependent variable in the dimension occupation (H4) differs from the above-mentioned procedure and is therefore explained at the end of this section.

Age deviance

For the dimension age (H1), the deviation of the actual portfolio of a backer from the expected portfolio is measured using a continuous dependent variable that reflects the extent of deviation (i.e., the delta) of the actual portfolio from the baseline. Thus, the continuous dependent variable (*Age deviance*) measures the extent to which the actual average age of the entrepreneurs in an investors' portfolio deviates from the expected average age of the entrepreneurs (the baseline)[2]. However, it should be noted that the baseline is not calculated simply based on all investments in the entire universe (as in the simplified example mentioned at the beginning of this chapter). Instead, the baseline is refined considering differences in the age of the entrepreneurs depending on the categories of the campaigns funded by an individual investor. This refinement is necessary because the distribution of entrepreneur characteristics differs across the different categories in which a project can be launched on Indiegogo. As shown in Appendix 5, the average age of the entrepreneurs funded in *Energy & Greentech* is almost 40 years. In comparison, the average age of entrepreneurs in the category *Fashion & Wearables* is only 29 years. Thus, to construct a baseline that reflects random choice as realistically as possible, it is necessary to calculate the baseline not only based on the average age of all entrepreneurs in the sample but to refine it depending on the average age of the entrepreneurs in the categories funded by the backer. Using my dataset, which includes the profile pictures of the entrepreneurs scraped from their Indiegogo account, I used the Python framework DeepFace[3] to estimate the age of the entrepreneurs based on their profile pictures.

Sex deviance and Race deviance

In the dimension sex (H2), the deviation of the actual portfolio of a backer from the expected portfolio is measured using a dummy variable (*Sex deviance*) to denote whether the actual share of the female sex in an investor's portfolio is above the expected share. The dependent variable is coded as 1 if the share of backings into female entrepreneurs is above the baseline and 0 otherwise. Something similar but slightly modified also applies to the dimension race (H3). In this dimension, I use a categorical variable (*Race deviance*), which I operationalize via six dummy variables—i.e., one dummy per race—to denote the deviation of a backer's actual portfolio from the baseline. The respective dummy is coded as 1 if an investor's actual share of backings into a particular racial group (such as black entrepreneurs) is above the expected value and 0 otherwise.

In both cases, the actual share of backings per group is measured as the relative frequency of a particular group of entrepreneurs (e.g., female or black entrepreneurs) among all investments of a backer. On the other hand, the baseline describes the expected share of backings per group, considering the differences among the various categories funded by an individual backer. Refining the baseline again makes sense because, as seen in Appendix 5, 32% of all backings in the category *Wellness* are pledged to female entrepreneurs. In comparison, only 3% of all investments in the category *Productivity* are received by female entrepreneurs. The same holds for the dimension race, where, for example, Asian entrepreneurs receive 19% of all investments in the category *Phones & Accessories*. In comparison, Asian entrepreneurs receive only 3% of all backings in the category *Comics*. Just as for the dimension age, I used the Python framework DeepFace to estimate both the sex and the race of the entrepreneurs based on their Indiegogo profile pictures. Additionally, I used the tool genderize.io[4] to determine the sex of the entrepreneurs based on their first names.

[2]The value of the dependent variable *Age deviance* is negative if the actual average age of the entrepreneurs in a portfolio is below the expected average age for the portfolio (indicating that the investor supports younger en-trepreneurs than expected) and vice versa.

[3]DeepFace is a lightweight face recognition and facial attribute analysis framework for Python. The facial attribute analysis allows to determine the age (in integers), sex (female or male) and race (black, white, Asian, Middle Eastern, Indian and Latino Hispanic) of an individual (Python Package Index, 2021a).

[4]Genderize.io is a tool to determine the sex (female or male) of a person given their first name. The tool offers an easy integration for all major languages—such as Python—and compares names with a database of more than 80.000 distinct names to determine the sex of an individual (Genderize.io, 2021).

Actual average distance and Expected average distance (baseline)
Despite using 'deviance' as the dependent variable in the previous dimensions, I use the two measures *Actual average distance and Expected average distance (baseline)* as the dependent variables in the dimension location (H5), to determine whether investors choose ties randomly or tend to invest in geographically close entrepreneurs. Thereby, comparable to what has already been presented above, the actual average distance denotes the average distance between an investor and all the entrepreneurs in the investor's portfolio, and the baseline reflects the expected average distance between an investor and the entrepreneurs, depending on the categories of the projects funded by the investor. Just as for the previous dimensions, refining the baseline considering the differences in the various categories supported by an investor makes sense. That is because Appendix 5 indicates that the average distance between backers and entrepreneurs in the category *Phones & Accessories* is about 6,756 kilometers. In comparison, the average distance between backers and entrepreneurs in the category *Dance & Theater* is only about 1,093 kilometers. To calculate the distances between the backers and entrepreneurs, I first converted the locations of all individuals (which I collected from the Indiegogo profile pages of the backers and founders) into geographic coordinates using the Python framework geopy[5]. Afterward, based on those coordinates, I calculated each dyad's distance (in kilometers) using the Haversine formula[6].

Funded occupation
Finally, it is worth mentioning a difference in the construction of the dependent variable used to analyze homophily in the dimension occupation (H4). While comparing the actual backings to the expected backings in the dependent variables of the previous dimensions, in this dimension it suffices to determine the top-level project category which was funded most frequently by a backer. Hence, I use a categorical dependent variable (*Funded occupation*), which I operationalize via three dummy variables—one dummy per top-level category—to denote the category in which a backer invests most frequently. The respective dummy is coded as 1 if this is the top-level category the investor invests most often into and 0 otherwise. For example, an investor has made five investments in the past, three in projects from the top-level category Tech & Innovation and two in projects from the top-level category Creative Works. The dependent variable would thus denote that the backer invests primarily into campaigns assigned to the top-level category Tech & Innovation, such that the respective dummy is coded as 1 and the other dummies are coded as 0. If no category was uniquely identified as most frequent among the backings of an investor, the backer was removed as an observation from the sample.

[5]Geopy is a Python framework for several popular geocoding services. The framework allows to locate the coor-dinates of addresses, cities, and countries across the globe (Python Package Index, 2021b).

[6]The haversine formula is used to calculate the distance between two points on Earth using their latitude and longitude (Python Package Index, 2021c).

3.3.2. Independent Variables

The various sociodemographic characteristics of the investors make up the key independent variables of my analyses.

Backer age
Since the profile pages on Indiegogo do not include information about an individual's birthday, I also used the facial attribute analysis tool DeepFace to infer the age of the backers from their Indiegogo profile pictures. Thus, to determine whether possible deviations in the investor's actual portfolio from the expected portfolio are due to the investor's preference for similar-age dyads, I use the independent variable (*Backer age*) to denote the age estimate of the investors in whole years.

Backer sex
I also derived data on the sex of the backers by analyzing their profile pictures, using the Python framework DeepFace. Additionally, as for the corresponding dependent variable, I also determined the sex of the investors using the genderize.io tool by algorithmically comparing their first names to the names from the genderize.io database. Thus, to test whether possible deviations in the actual backer portfolio from the baseline are due to a preference for same-sex dyads, I use an independent dummy variable (*Backer sex*) to indicate the sex estimate, where male represents the omitted baseline sex. Thus, the variable is coded as 1 if the investor is female and 0 otherwise.

Backer race
Just as for the previous independent variables, I also derived data on the race of the backers by analyzing their Indiegogo profile pictures using the Python framework DeepFace. I use a categorical independent variable (*Backer race*), which I operationalize via multiple dummy variables to denote an investor's estimated belonging to a particular racial group (i.e., black, white, Asian, Indian, Middle Eastern, or Latino Hispanic). The respective dummy is coded as 1 if the backer belongs to the specific racial group and 0 otherwise, where in general, white represents the omitted baseline race. However, please note that this differs for the various models to test for racial homophily (H3), where the white race, depending on the model, is also part of the analysis.

Backer occupation
To test whether investors show a tendency to fund projects from entrepreneurs with the same occupation as themselves (approximated by the top-level category in my study), I use a categorical independent variable (*Backer occupation*). This variable denotes the top-level category in which the backer is mainly active, i.e., the top-level category in which the backer has launched most own campaigns. For example, suppose a backer has launched three campaigns. Two in Tech & Innovation and one in Creative Works. In that case, the independent variable *Backer occupation* denotes that Tech & Innovation is the category in which the backer mainly launches own campaigns. I used the data about all backers in my dataset who eventually became entrepreneurs on Indiegogo (i.e., those backers who have also launched at least one cam-

paign) to determine the top-level categories in which those entrepreneur-backers started most of their projects. Like for the corresponding dependent variable, those backers were removed from the sample, where no category was identified as most frequent among the campaigns they launched.

3.3.3. Control Variables

Following what prior research in the context of online resource acquisition presented as influential concerning the entrepreneurs in crowdfunding, I include several backer-related variables in my analysis to control for differences in their experience and the quality of the backer profiles. Additionally, my study also controls for the influence of the categories funded by the backers and their sociodemographic characteristics such as age, sex, and race.

First, I control for backer experience with several discrete variables. Following previous studies in the context of crowdfunding (see Groza et al. (2020); Taeuscher, Bouncken, and Pesch (2020); Tan and Reddy (2021)), I control for backer experience by measuring the number of investments made by a backer on Indiegogo (*Number of contributions*). In addition, I control for the number of projects launched by the backer (*Number of own projects*), to control for differences in crowdfunding experience (compare Allison, Davis, Webb, and Short (2017); Cornelius and Gokpinar (2020); Soublière and Gehman (2020); Taeuscher et al. (2020); Tan and Reddy (2021)). Finally, I control for the number of comments a backer has made on the platform (*Number of comments*) to account for differences in platform activity.

Second, I use several control variables to account for differences in the quality of the backer profiles. Because users can link their Facebook account to Indiegogo, it is possible to determine the number of friends of an individual. Thus, I use a discrete variable to denote the number of Facebook friends of a backer (*Number of FBF*), which has previously been used as a potential indicator of the size of an individual's network (exemplary see Anglin, Wolfe, Short, McKenny, and Pidduck (2018); Mollick (2014); Younkin and Kuppuswamy (2018)). Similar to previous studies focusing on the campaign and entrepreneur (see Davis, Hmieleski, Webb, and Coombs (2017); Taeuscher et al. (2020); Younkin and Kuppuswamy (2018)), I control for the quality of the backer profiles using a discrete variable that reflects the number of words a backer uses on the profile page (*Profile word count*). This measure indicates the differences that backers put into creating and maintaining their Indiegogo profile page.

Furthermore, I use a dummy variable to control for the account quality of the backers, indicating whether the Indiegogo account of a backer is verified (*Profile verified*), where the dummy variable is coded as 1 if the account is verified through at least one channel, such as email, and 0 otherwise. Also, comparable to what Scheaf et al. (2018) did at the entrepreneur level, I use dummies to indicate whether a backer has its Facebook account connected to Indiegogo (*Facebook connected*) and whether a backer has its Twitter account linked (*Twitter connected*). For both variables, the dummy is coded as 1 if the backer has the respective account connected and 0 otherwise.

Third, as already explained, projects on Indiegogo are classified according to different categories. Consistent with previous research (see Allison et al. (2017); Cornelius and Gokpinar (2020); Oo, Allison, Sahaym, and Juasrikul (2019); Taeuscher et al. (2020)) and to control for heterogeneity across the various categories, I control for the category in which the backer invests most often. Hence, I include a categorical control variable (*Main category funded*) to denote the category the backer supports most often. Please note that the dependent variable *Funded occupation* (which is used for the analysis in the dimension occupation—see chapter 3.3.1) in principle measures the same. It does so, however, at the top-level category layer. The control variable *Main category funded*, on the other hand, controls at the more granular category layer and is not used in the models to analyze the dimension occupation. Those backers where no category was identified as most frequent among their investments were removed from the sample.

Finally, the respective sociodemographic characteristics used as independent variables (i.e., the age, sex, and race of the backer—see chapter 3.3.2) are vice versa also used as control variables.

3.4. Descriptive Statistics

To provide a comprehensive overview of the dataset, the chapter is structured as follows. First, I present the descriptive statistics based on the overall dataset of roughly 3.5 million backings. This includes, among others, the correlation matrix of the various variables used in my analyses, as well as the baseline table containing the different entrepreneur characteristics per Indiegogo category (which I calculated and used to determine the expected backer portfolios in my analyses). Next follow some descriptive statistics regarding the set of unique backers in my dataset (including, for example, the distribution of sex and race among the backers as well as additional summary statistics). Afterward, the set of unique entrepreneurs in the dataset is briefly described as well. Fourth, I provide a superficial description of the individual campaigns in my dataset, which mainly provides a first impression about the projects funded, including the amounts of money pledged, the project duration, and the like. Finally, and particularly relevant for the analysis in the dimension occupation, this chapter ends with a brief description of specifically those backers who themselves act as entrepreneurs on Indiegogo.

Appendix 2 includes the correlation matrix based on the overall dataset, covering the numerical and dummy variables used in my regression models and provides an overview of the interactions between the various variables. Furthermore, Appendix 3 includes the summary statistics of those variables based on the overall dataset. The statistics show, for example, that among all investments, the average age of the backers is almost 32 years, and the average distance between an investor and the entrepreneurs funded is about 5,045 kilometers. Additionally, table 2 contains an overview of the distribution of sex and race among the backers and entrepreneurs based on all investments.

The table shows that female investors make about 28% of all backings (considering the numbers from genderize.io). On the other side of the dyad, female entrepreneurs receive roughly 22% of all investments (considering the numbers from genderize.io). Furthermore, most investments are made by white backers, making up 68% of all backings. Likewise, white entrepreneurs receive most investments, accounting for 73% of all investments on the receiver side. However, please note that the N for the various tables in this chapter varies depending on whether the specific characteristics were identified for the individuals.

In addition, Appendix 4 provides an overview of the distribution of the roughly 3.5 million investments across the different categories on Indiegogo. Thereby, most backings are pledged to campaigns assigned to the category Film (16% of all investments), followed by Phones & Accessories (12% of all investments), and Health & Fitness (7% of all investments), with two of those most frequently funded categories belonging to the top-level category Tech & Innovation. In the context of different categories, it is worth mentioning the table I calculated and used to determine the different baselines (i.e., the expected portfolios of the individual backers). This table (see Appendix 5) contains the distribution of entrepreneur characteristics, such as their average age or the share per racial group, among all investments in my dataset aggregated per Indiegogo category. Here, its main purpose is to provide an overview of differences among the entrepreneur characteristics in the various categories offered by Indiegogo.

While the previous descriptions were based on all investments, the unique backers in my dataset will now be described. Table 3 provides an overview of some of the main variables describing the backers. The statistics show that the average investor is 31.36 years old and makes just under four contributions on the platform. In addition, an average backer leaves only roughly one comment on the platform and is hardly active as an entrepreneur on Indiegogo, as the average number of own projects of a backer is only 0.03.

Table 4 furthermore provides an overview of the distribution of sex and race among the unique backer population. Interestingly, while according to the numbers from genderize.io, 32% of all backers are female (compare table 4), they account for only 28% of all investments (compare table 2). This pattern suggests that, on average, female investors make fewer investments than men. The same pattern occurs considering the numbers from DeepFace. Furthermore, only 0.7% of the roughly 1.2 million different backers have their Facebook profile connected to their Indiegogo account, and only 0.6% of the investors have their Twitter profile linked. However, about 24% of all backer profiles are verified via at least one channel, such as email.

Although the entrepreneurs are not the focus of my study, some descriptive statistics will nevertheless be presented to provide a complete overview of the dataset. Table 5 includes a summary describing the unique founders. This table indicates that the average entrepreneur is 30.32 years old and makes investments on the platform as well (2.19 contributions on average), but less than the backers (with an average of 3.86 contributions, as presented in table 3). However, with an average of almost two comments, entrepreneurs are more active than backers in this respect. Furthermore, an average entrepreneur launches 1.30 campaigns on Indiegogo.

Table 6 furthermore provides an overview of the distribution of sex and race among the unique entrepreneurs. Comparing those numbers to table 4, it becomes clear that the distribution of the characteristics among the unique entrepreneurs is very similar to the distribution among the unique backers. For example, 69% of all entrepreneurs (compare table 6) and 67% of all backers (compare table 4) are white.

Furthermore, the unique campaigns in my dataset are briefly described in the following. However, please note that this is purely for an initial overview of the dataset. None of these variables are included in my analyses, as the focus of my study is on the backers in crowdfunding. Table 7 summarizes some essential variables describing the campaigns funded. Please be aware that, for the sake of simplicity, table 7 only includes projects launched in the US$ currency, which make up by far the largest share of projects in my dataset.

On average, a campaign is backed by almost 111 investors, has 4.34 updates posted, and contains roughly 24 comments. Interestingly, the average project duration in my dataset reflects the recommendation made by Indiegogo. As already presented in chapter 3.1, Indiegogo recommends a campaign duration of about 40 days. My dataset indicates comparable numbers. Although campaigns on Indiegogo can, in principle, last up to 60 days, the average duration in my dataset is 42.51 days, which is close to the recommendation of Indiegogo.

Furthermore, as described in chapter 3.1, campaigns on Indiegogo can either have a flexible goal or a fixed goal. The vast majority of projects in my dataset have a flexible goal (i.e., 97% of all projects), whereas only 3% of the campaigns in my dataset have a fixed goal. Moreover, 60% of all projects in my sample are assigned to the top-level category Creative Works, followed by Community Projects, which account for 21% of all campaigns. Only 19% of all projects are launched in the top-level category Tech & Innovation. This is interesting because, as presented at the beginning of this chapter, projects from Tech & Innovation are among the most frequently funded projects in my dataset, which indicates that these campaigns receive an above-average number of investments. Appendix 6 contains the fine-grained share of campaigns launched per category. Furthermore, it should also be noted that—as presented in chapter 3.1—Indiegogo states that women launch 47% of campaigns that exceed their funding goal. My sample shows a somewhat similar pattern, where women initiated 37% of the campaigns exceeding their goal.

Finally, those investors who act as entrepreneurs on the platform Indiegogo themselves should briefly be mentioned,

Table 2: Distribution of sex and race among backers and entrepreneurs (all investments)

Source: Own illustration

Characteristic	Share of all investments made	Share of all investments received
Female (DeepFace)	16%	11%
Male (DeepFace)	84%	89%
Female (genderize.io)	28%	22%
Male (genderize.io)	72%	78%
White	68%	73%
Asian	11%	10%
Latino Hispanic	7%	5%
Black	6%	5%
Middle Eastern	6%	6%
Indian	2%	1%

Table 3: Summary statistics of the unique backers

Source: Own illustration

Variable	Min.	Max.	Median	Mean	SD
Backer age	20	67	31	31.36	5.90
Number of contributions	1	3,802	2	3.86	8.79
Number of comments	0	2,003	0	0.85	7.98
Number of own projects	0	235	0	0.03	0.31
Number of FBF	0	5,034	165	371.57	598.69
Profile word count	0	1,548	0	0.49	8.94

Table 4: Distribution of sex and race among the unique backers

Source: Own illustration

Characteristic	Share of Backers
Female (DeepFace)	17%
Male (DeepFace)	83%
Female (genderize.io)	32%
Male (genderize.io)	68%
White	67%
Asian	12%
Latino Hispanic	7%
Middle Eastern	6%
Black	6%
Indian	2%

Table 5: Summary statistics of the unique entrepreneurs

Source: Own illustration

Variable	Min.	Max.	Median	Mean	SD
Founder age	17	70	30	30.32	5.33
Number of contributions	0	1,600	1	2.19	8.36
Number of comments	0	3,016	0	1.97	19.42
Number of own projects	1	228	1	1.30	1.25

Table 6: Distribution of sex and race among the unique entrepreneurs

Source: Own illustration

Characteristic	Share of Founders
Female (DeepFace)	20%
Male (DeepFace)	80%
Female (genderize.io)	38%
Male (genderize.io)	62%
White	69%
Asian	9%
Latino Hispanic	6%
Middle Eastern	6%
Black	8%
Indian	2%

Table 7: Summary statistics of unique campaigns funded in US$

Source: Own illustration

Variable	Min.	Max.	Median	Mean	SD
Number of backers	0	104,448	29	110.71	863.88
Number of updates	0	345	1	4.34	8.61
Number of comments	0	28,210	5	24.18	230.32
Campaign duration (days)	0	2,000	40	42.51	24.13
Total funding (US$)	0	17,595,711	1,911	17,192.46	155,797.84
Campaign start date	2008-01-14	2021-02-05	-	2015-04-17	-
Campaign end date	2010-04-17	2021-04-05	-	2015-06-14	-

which is particularly interesting for my analysis in the dimension occupation. Among the roughly 1.2 million different backers in my dataset, 27,916 have started at least one campaign on Indiegogo. These entrepreneur-backers launch most projects in the top-level category Creative Works (accounting for 69% of all projects launched by them), followed by the top-level category Tech & Innovation (17% of these investors' projects) and Community Projects (14% of all entrepreneur-investor projects). Considering the distribution among the more fine-grained categories, entrepreneur-backers launch most projects in the category Film (35% of all projects), followed by the category Dance & Theater (13% of all projects). These two categories, both from the top-level category Creative Works, therefore account for roughly 48% of all projects launched by these investors. On average, an entrepreneur-backer is 31 years old and is more active than a standard backer, as this person on average launches 1.25 campaigns and contributes to projects 4.81 times. Interestingly, in contrast to what has been presented so far, this investor type shows a more even distribution of sex in my dataset. Just under 42% of all entrepreneur-investors are women, and about 58% are men (considering the numbers from genderize.io). However, white entrepreneur-investors make up the majority here as well, as they account for almost 68% of these investors, followed by Asian entrepreneur-investors, who make up just under 10%.

3.5. Methods

I conducted different quantitative analyses to test the hypotheses I developed throughout chapter 2.3. The differences in the analytical methods used to test my hypotheses arise from the differences in the type and calculation of the dependent variables. I performed regression analyses to study homophily in the dimensions *age*, *sex*, *race*, and *occupation*, and I performed a t-test to analyze homophily in the dimension *location*. In the following, the methods used to test my hypotheses are described in more detail.

In particular, to analyze homophily in the dimension age (H1), I ran an ordinary least squares regression including robust standard errors to tackle heteroscedasticity. The model I ran in this context measures the interaction between the independent variable *Backer age* and the dependent variable *Age deviance* to determine whether investors tend to invest in entrepreneurs they are similar to with respect to age. I used the variance inflation factor to tackle multicollinearity, with the factor being well below 1.7 for all non-categorical variables. Please note that the variance inflation factor was also calculated for all other models in my analyses, whereby the factor is always well below 1.7.

To analyze homophily in the dimensions sex (H2), race (H3), and occupation (H4), I ran logistic regression models. For the analysis in the dimension sex, I ran one logistic regression model. This model measures the interaction between the independent variable *Backer sex* (i.e., the female sex) and the dependent variable *Sex deviance* (also for the female sex) to investigate whether investors tend to support more entrepreneurs of their sex than suggested by chance. Since the characteristic sex can only be male or female, one model is sufficient to make conclusions for both women and men regarding homophilous choices.

For the analysis in the dimension race, I ran six models (model 3a to 3f), that is, one model per race. In each model, I measure the interactions of the independent variable *Backer race* (i.e., every race apart from the one tested for in the dependent variable) and the respective dependent variable *Race deviance*. The model for the white race, for example, tests whether all races apart from the group of white investors discriminate against white entrepreneurs. If this is true, I could conversely infer that the omitted group of white investors supports more white entrepreneurs than expected, ultimately indicating homophily.

For the analysis in the dimension occupation, I ran three logit models (model 4a to 4c), that is, one model per top-level category. Thereby, I measure the interaction between the categorical independent variable *Backer occupation* and one expression of the dependent variable *Funded occupation* (i.e., Creative Works, Tech & Innovation, or Community Projects). These models are used to analyze whether investors tend to invest mainly in entrepreneurs who are active in the same category as themselves and thus share the same occupation for the sake of this study.

Unlike the previous dimensions, I conducted a one-sided t-test in the dimension location (H5) and compared the *Actual average distance* with the *Expected average distance* to analyze whether backers tend to invest in entrepreneurs whom they are geographically closer to than chance would suggest.

4. Results

The following chapter covers the results of my analyses. Chapter 4.1 presents the results of the models testing hypotheses 1 to 5. Subsequently, chapter 4.2 presents the results of the robustness test. Specifically, this includes the analysis in the dimension sex (H2) using the same model as in chapter 4.1, but with the alternative data from the tool genderize.io to validate the results of the corresponding model in the previous chapter.

4.1. Hypotheses Testing

Dimension Age (H1)

In hypothesis 1, I predicted that contributions to entrepreneurs with a similar age as the investor occur more frequently than we would expect if investors chose projects for contribution at random. That is, investors make homophilous choices as they tend to invest in entrepreneurs whom they are similar to in age. Thus, I predicted a positive relationship between the age of the backer and the age deviance.

Figure 1 presents the result of the ordinary least squares regression model used to test the hypothesis. The results support H1, as the relationship between the age of the backer and the age deviance—indicating how far the actual average age of the entrepreneurs funded by the backer deviates from the expected average age—is positive and significant ($b = 0.10$, $p < .001$). This indicates that for each additional year of backer age, the age deviance also increases by 0.10. The results suggest that, for example, older investors tend to support older entrepreneurs than suggested by chance, thus indicating homophily.

Dimension Sex (H2)

In hypothesis 2, I predicted that contributions to entrepreneurs of the same sex as the investor occur more frequently than we would expect if investors chose projects for contribution at random. Thus, I predicted that backers tend to invest in entrepreneurs with whom they share the same sex. That is, they make homophilous choices with respect to the sociodemographic dimension sex.

Figure 2 presents the result from the logit regression model used to test this hypothesis. The results support H2, as the interaction of the backer-sex with the corresponding sex deviance is positive and significant. Particularly for the model in figure 2, the interaction of the female backer-sex with the female sex deviance—indicating whether a backer supports more female entrepreneurs than expected—is positive and significant ($b = 0.89$, $p < .001$). This result indicates that for female investors, compared to male investors, the odds to fund more female entrepreneurs than expected increase by a factor of 2.43. Since the sociodemographic characteristic sex has only two forms—male and female—it can further be inferred from the results that male investors also tend to support more male entrepreneurs than chance would suggest. Therefore, the results demonstrate that investors, in general, tend to invest in entrepreneurs of the same sex, indicating homophily with respect to this sociodemographic dimension.

Dimension Race (H3)

In hypothesis 3, I predicted that contributions to entrepreneurs of the same race as the investor occur more frequently than we would expect if investors chose projects for contribution at random. Thus, I predicted that investors tend to invest in entrepreneurs with whom they share the same race. That is, they make homophilous choices in the sociodemographic dimension race.

Figures 3 to 8 show the logit regression models used to test for homophily in this dimension. First, the results for the black race are presented (model 3a), followed by the results for the Indian race (model 3b), white race (model 3c), Asian race (model 3d), Latino Hispanic race (model 3e) and finally the results for the Middle Eastern race (model 3f). However, it should already be mentioned that I find mixed results, and therefore, overall, hypothesis 3 is not supported, as significant differences can be observed depending on the specific race considered.

Variables	b	Std Err	p
Independent Variable			
Backer age	0.10***	0.02	0.00
Control Variables			
Backer sex	0.29	0.32	0.37
Backer race		Included	
Main category funded		Included	
Number of own projects	0.01	0.12	0.95
Number of comments	-0.01	0.01	0.39
Number of contributions	0.00	0.01	0.94
Number of FBF	0.00	0.00	0.96
Facebook connected	0.33	0.33	0.32
Twitter connected	-0.33	0.36	0.37
Profile verified	0.29	0.26	0.26
Profile word count	-0.01*	0.00	0.02

*Notes. N = 1,246; ***p < .001; **p < .01; *p < .05; † p < .1*

Figure 1: Ordinary least squares model for the dimension age

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer sex (Female)	0.89***	0.17	0.00
Control Variables			
Backer age	0.02†	0.01	0.07
Backer race		Included	
Main category funded		Included	
Number of own projects	0.09	0.07	0.22
Number of comments	-0.01	0.01	0.41
Number of contributions	0.02***	0.00	0.00
Number of FBF	0.00	0.00	0.86
Facebook connected	0.08	0.19	0.67
Twitter connected	0.03	0.20	0.87
Profile verified	-0.09	0.15	0.56
Profile word count	0.00	0.00	0.27

*Notes. N = 1,246; ***p < .001; **p < .01; *p < .05; † p < .1*

Figure 2: Logit model for the dimension sex

Source: Own illustration

Model 3a: Black Race
As shown in figure 3, considering the black race deviance—indicating whether a backer supports more black entrepreneurs than expected—the interaction is negative and significant with all backer races except the omitted group of black investors, hence indicating homophilous choices made by black investors. Specifically for white backers, the interaction is negative and significant (b = -1.59, p < .001). Thus, for white investors, compared to black investors, the odds to fund more black entrepreneurs than expected change by a factor of 0.20. The same is true for Asian investors, where the interaction is also negative and significant (b = -1.15, p < .01). Therefore, for Asian investors, the odds to fund more black entrepreneurs than expected change by a factor of 0.32, which means that being an Asian investor, the probability to fund more black entrepreneurs than expected decreases significantly compared to being a black investor. Also, for the Indian backer group, the interaction is nega-

tive and significant (b = -2.68, p < .05). Thus, for Indian investors, the odds to fund more black entrepreneurs than expected change by a factor of 0.07, compared to black investors. Similar holds for Latino Hispanic investors, where interaction is negative and significant (b = -2.10, p < .001). Thus, for Latino Hispanic investors, the odds to fund more black entrepreneurs than expected change by a factor of 0.12, indicating that the odds to fund more black entrepreneurs than expected significantly decrease if the investor is Latino Hispanic instead of the investor being black. Finally, for the Middle Eastern investors, the interaction is also negative and significant (b = -1.84, p < .001). Therefore, for Middle Eastern investors, the odds to fund more black entrepreneurs than expected changes by a factor of 0.16, compared to the omitted group of black investors.

Since the interaction is negative and significant for all five investor groups in this model, indicating that they all significantly discriminate against black entrepreneurs, it can be conversely concluded that the interaction with the omitted group of black investors is positive and significant. Thus, black investors tend to support more black entrepreneurs than chance would suggest, arguing for homophilous choices. Therefore, considering this model only, the results would support hypothesis 3.

Model 3b: Indian Race

Similar results are found for Indian backers. As shown in figure 4, considering the Indian race deviance—indicating whether an investor supports more Indian entrepreneurs than expected—the interaction with the various racial groups of the investors used in this model is negative and significant. Specifically, for the group of white backers, the interaction is negative and significant (b = -2.56, p < .001). Thus, for white investors compared to Indian investors, the odds of funding more Indian entrepreneurs than expected change by a factor of 0.08, indicating that the probability significantly decreases for white backers. The same is true for black investors, where the interaction is also negative and significant (b = -1.84, p < .05). Thus, for black investors, the odds of funding more Indian entrepreneurs than chance would suggest change by a factor of 0.16, compared to Indian investors. Likewise, for the Asian backer group, the interaction is negative and significant (b = -2.24, p < .01). Therefore, for Asian investors compared to Indian investors, the odds of funding more Indian entrepreneurs than expected change by a factor of 0.11, indicating a decreasing probability. Also, for Latino Hispanic investors, the interaction is negative and significant (b = -2.94, p < .05). Therefore, the odds of funding more Indian entrepreneurs than expected change by a factor of 0.05 if the investor is Latino Hispanic, indicating that the likelihood to fund more Indian entrepreneurs than expected decreases drastically if the investor is Latino Hispanic instead of Indian. Finally, for Middle Eastern backers, the interaction is also negative and significant (b = -2.15, p < .05). Thus, for Middle Eastern investors, the odds of funding more Indian entrepreneurs than expected change by a factor of 0.12, compared to the omitted group of Indian investors.

Since the interaction is negative and significant for all five backer races in this model, it can conversely be inferred that the interaction with the omitted group of Indian investors is positive and significant. That is, Indian investors tend to support significantly more Indian entrepreneurs than expected from random choice. Thus, considering only this model, I find support for my hypothesis on racial homophily.

Model 3c: White Race

Other than what has been presented before, the situation is different for white backers. As shown in figure 5, considering the white race deviance, the interaction with the five backer races included in this model (i.e., every race apart from the group of white investors) is negative, yet not always significant. Thereby, for the group of black backers, the interaction is negative and significant (b = -1.30, p < .001). Thus, for black investors compared to white investors, the odds to fund more white entrepreneurs than expected change by a factor of 0.27, indicating a decreasing likelihood of supporting more white entrepreneurs than expected if the backer is black. The same holds for Asian investors, where the interaction is negative and significant (b = -0.70, p < .001). Thus, for Asian investors, the odds to fund more white entrepreneurs than expected change by a factor of 0.50, meaning that for Asian investors, the odds to fund more white entrepreneurs than expected are only half as high as for white investors. The interaction is also negative and significant for Latino Hispanic backers (b = -0.74, p < .01). Thus, for Latino Hispanic investors, the odds to fund more white entrepreneurs than expected change by a factor of 0.48, indicating a decreasing probability compared to white investors. On the other hand, for Indian investors, the interaction is negative but not significant (b = -0.66, p < .10). Likewise, for Middle Eastern investors, the interaction is also negative but not significant (b = -0.34, p = .19).

Although the interaction is negative for all five investor groups in this model, suggesting that white entrepreneurs tend to make homophilous choices, I still cannot support hypothesis 3 considering only this model. This is because some interactions with the independent variable are not significant in this model. Consequently, I cannot conversely conclude that, in general, white backers support more white entrepreneurs than expected.

Model 3d: Asian Race

As shown in figure 6, considering the Asian race deviance—indicating whether an investor supports more Asian entrepreneurs than expected—the interaction with the different racial groups of investors in this model is negative, yet not always significant. Thereby, for white backers, the interaction is negative and significant (b = -0.78, p < .001). Thus, for white investors compared to Asian investors, the odds of funding more Asian entrepreneurs than expected change by a factor of 0.46, meaning that for a white investor, the likelihood of supporting more Asian entrepreneurs than chance would suggest decreases. The same is true for Middle Eastern investors, where the interaction is negative and

Variables	b	Std Err	p
Independent Variable			
Backer race (White)	-1.59***	0.28	0.00
Backer race (Asian)	-1.15**	0.37	0.00
Backer race (Indian)	-2.68*	1.08	0.01
Backer race (Latino Hispanic)	-2.10***	0.54	0.00
Backer race (Middle Eastern)	-1.84***	0.53	0.00
Control Variables			
Backer sex	0.28	0.24	0.24
Backer age	0.00	0.02	0.79
Main category funded		Included	
Number of own projects	0.07	0.09	0.43
Number of comments	0.00	0.01	0.99
Number of contributions	0.01**	0.01	0.01
Number of FBF	0.00***	0.00	0.00
Facebook connected	-0.33	0.26	0.20
Twitter connected	0.13	0.27	0.62
Profile verified	0.07	0.20	0.71
Profile word count	0.00	0.00	0.90

Notes. $N = 1{,}246$; *** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .1$

Figure 3: Logit model 3a for the dimension race (Race deviance—Black)

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer race (White)	-2.56***	0.67	0.00
Backer race (Black)	-1.84*	0.86	0.03
Backer race (Asian)	-2.24**	0.80	0.01
Backer race (Latino Hispanic)	-2.94*	1.19	0.01
Backer race (Middle Eastern)	-2.15*	0.92	0.02
Control Variables			
Backer sex	0.73†	0.43	0.09
Backer age	0.03	0.03	0.35
Main category funded		Included	
Number of own projects	-0.06	0.23	0.80
Number of comments	-0.01	0.02	0.55
Number of contributions	0.02**	0.01	0.01
Number of FBF	0.00*	0.00	0.02
Facebook connected	-0.39	0.52	0.45
Twitter connected	-0.15	0.56	0.79
Profile verified	-0.18	0.40	0.66
Profile word count	0.00	0.01	0.94

Notes. $N = 1{,}246$; *** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .1$

Figure 4: Logit model 3b for the dimension race (Race deviance—Indian)

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer race (Black)	-1.30***	0.25	0.00
Backer race (Asian)	-0.70***	0.20	0.00
Backer race (Latino Hispanic)	-0.74**	0.25	0.00
Backer race (Indian)	-0.66†	0.39	0.09
Backer race (Middle Eastern)	-0.34	0.26	0.19
Control Variables			
Backer sex	-0.34*	0.16	0.03
Backer age	0.00	0.01	0.91
Main category funded		Included	
Number of own projects	0.11	0.07	0.11
Number of comments	0.00	0.00	0.26
Number of contributions	-0.01*	0.00	0.02
Number of FBF	0.00**	0.00	0.01
Facebook connected	0.06	0.16	0.71
Twitter connected	0.07	0.17	0.68
Profile verified	-0.11	0.13	0.41
Profile word count	0.00	0.00	0.56

*Notes. N = 1,246; ***p < .001; **p < .01; *p < .05; † p < .1*

Figure 5: Logit model 3c for the dimension race (Race deviance—White)

Source: Own illustration

significant (b = -0.84, p < .05). Thus, for Middle Eastern investors, the odds of funding more Asian entrepreneurs than expected change by a factor of 0.43, indicating that for Middle Eastern investors, the probability of funding more Asian entrepreneurs than expected is less than half of that for the omitted group of Asian investors. On the other hand, the results for the other three backer races included in this model are negative but not significant. For the group of black investors, the interaction is negative but not significant (b = -0.61, p < .10). Likewise, for Indian backers, the interaction is negative, yet not significant (b = -0.28, p = .57). Furthermore, the interaction for the Latino Hispanic investor group is negative but insignificant as well (b = -0.08, p = .82).

Again, the results of this model do not support hypothesis 3. Although the interaction is negative for all five investor races in this model, which would suggest that the interaction of the omitted group of Asian investors with the Asian race deviance is positive, this conclusion is not possible. That is because the interaction with three of five races is not significant. Thus, for the Asian race, it cannot be inferred that, on average Asian investors tend to support more Asian entrepreneurs than chance would suggest. Hence, I do not find support for my hypothesis on racial homophily considering this model only.

Model 3e: Latino Hispanic Race

As shown in figure 7, considering the Latino Hispanic race deviance, the interaction with the five investor races in the model is also negative but not always significant. For white investors, the interaction is negative and significant (b = -0.84, p < .01). Thus, for white backers compared to Latino Hispanic backers, the odds of funding more Latino Hispanic entrepreneurs than expected change by a factor of 0.43. The same is true for black backers, where the interaction is also negative and significant (b = -1.11, p < .05). Thus, for black investors, the odds of funding more Latino Hispanic entrepreneurs than expected change by a factor of 0.33, indicating that for black investors, the likelihood to fund more Latino Hispanic entrepreneurs than expected is only about one-third as high as for Latino Hispanic investors. On the other hand, the results for the three other races included are negative but not significant. For Asian investors, the interaction is negative but not significant (b = -0.69, p < .10). Likewise, for Indian investors, the interaction is also negative but insignificant (b = -0.92, p = .19). Furthermore, the interaction is negative yet not significant for the Middle Eastern investor group (b = -0.79, p < .10).

Therefore, similar to what has been presented for the white and Asian race (models 3c and 3d), the results of this model do not support hypothesis 3. This is because even though the interaction for all five investor races in this model is negative, indicating a positive interaction of the omitted group of Latino Hispanic investors with the Latino Hispanic race deviance, the interaction is not significant for three of the five races in this model. Thus, since not all racial groups included in this model significantly discriminate against Latino Hispanic entrepreneurs, it is not possible

Variables	b	Std Err	p
Independent Variable			
Backer race (White)	-0.78***	0.22	0.00
Backer race (Middle Eastern)	-0.84*	0.39	0.03
Backer race (Black)	-0.61†	0.36	0.10
Backer race (Indian)	-0.28	0.49	0.57
Backer race (Latino Hispanic)	-0.08	0.34	0.82
Control Variables			
Backer sex	0.26	0.20	0.19
Backer age	0.01	0.01	0.49
Main category funded		Included	
Number of own projects	-0.08	0.09	0.40
Number of comments	0.01†	0.00	0.08
Number of contributions	0.01*	0.00	0.01
Number of FBF	0.00	0.00	0.79
Facebook connected	0.05	0.20	0.81
Twitter connected	-0.17	0.22	0.44
Profile verified	0.17	0.16	0.30
Profile word count	0.00	0.00	0.65

*Notes. N = 1,246; ***p < .001; **p < .01; *p < .05; † p < .1*

Figure 6: Logit model 3d for the dimension race (Race deviance—Asian)

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer race (White)	-0.84**	0.32	0.01
Backer race (Black)	-1.11*	0.52	0.03
Backer race (Asian)	-0.69†	0.40	0.08
Backer race (Indian)	-0.92	0.70	0.19
Backer race (Middle Eastern)	-0.79†	0.48	0.10
Control Variables			
Backer sex	-0.08	0.25	0.74
Backer age	-0.01	0.02	0.58
Main category funded		Included	
Number of own projects	-0.16	0.12	0.18
Number of comments	0.00	0.00	0.59
Number of contributions	0.01†	0.00	0.06
Number of FBF	0.00	0.00	0.34
Facebook connected	0.59*	0.24	0.01
Twitter connected	0.02	0.25	0.95
Profile verified	-0.20	0.20	0.32
Profile word count	0.00	0.00	0.50

*Notes. N = 1,246; ***p < .001; **p < .01; *p < .05; † p < .1*

Figure 7: Logit model 3e for the dimension race (Race deviance—Latino Hispanic)

Source: Own illustration

to conversely conclude that the omitted group of Latino Hispanic investors exhibits a bias to support significantly more Latino Hispanic entrepreneurs than would be expected by chance.

Model 3f: Middle Eastern Race

As shown in figure 8, considering the Middle Eastern race deviance—indicating whether an investor supports more Middle Eastern entrepreneurs than expected—the interaction with the various racial groups of investors included in this model is negative but not significant. Specifically, for white investors, the interaction is negative but not significant ($b = -0.31$, $p = .38$). The same is true for black investors, where the interaction is negative but not significant ($b = -0.33$, $p = .51$). Likewise, for the group of Asian backers, the interaction is negative but insignificant ($b = -0.66$, $p = .16$). The same holds for Indian investors, where the interaction is also negative yet not significant ($b = -0.92$, $p = .27$). Finally, the interaction is negative but insignificant for the group of Latino Hispanic investor as well ($b = -0.88$, $p = .13$).

Since the results for each of the five backer races in this model are not significant, it is not possible to conclude that Middle Eastern entrepreneurs are significantly discriminated against by investors from all but their racial group. Thus, it is also impossible to conversely conclude that Middle Eastern investors make homophilous choices by supporting more Middle Eastern entrepreneurs than expected. Therefore, considering this model only, the results do not support hypothesis 3.

To summarize the analysis of hypothesis 3, I cannot conclude that investors in general tend to support more entrepreneurs of their race than expected. Thus, hypothesis 3 is not supported as I find mixed results with only models 3a (black) and 3b (Indian) supporting the hypothesis and models 3c to 3f not supporting the hypothesis.

Dimension Occupation (H4)

In hypothesis 4, I predicted that contributions to entrepreneurs with the same occupation as the investor occur more frequently than contributions to entrepreneurs with a different occupation than the investor. I predicted that investors who are active as entrepreneurs on Indiegogo tend to invest in entrepreneurs with whom they share the same occupation. That is, they invest in projects of entrepreneurs in the same top-level category as they launch their own projects.

Figures 9 to 11 present the results from the logit regression models used to test hypothesis 4. Each model corresponds to one top-level category on Indiegogo, representing the occupation in the context of this study. First, the results for the top-level category Creative Works are presented (model 4a), followed by the model for the top-level category Tech & Innovation (model 4b). Finally, the results for the top-level category Community Projects (model 4c) are presented. However, it should already be mentioned that the results of all three models support hypothesis 4, suggesting that investors make homophilous choices as they tend to mainly invest in entrepreneurs with whom they share the same occupation.

Model 4a: Funded Occupation—Creative Works

As shown in figure 9, considering Creative Works as the funded occupation in the dependent variable, the interaction with Creative Works as the backer occupation in the independent variable is positive and significant ($b = 3.42$, $p < .001$). This result suggests that for investors who are active in Creative Works, the odds of mainly funding entrepreneurs active in Creative Works increase by a factor of 30.62, compared to investors who are active in Community Projects. Thus, the result of this model supports H4, as investors significantly tend to invest in the same occupation as they are active when considering the top-level category Creative Works.

Model 4b: Funded Occupation—Tech & Innovation

Comparable results are found for the funded occupation Tech & Innovation (see figure 10). Also in this case, the interaction with the corresponding backer occupation is positive and significant ($b = 2.17$, $p < .001$). These results suggest that for investors who are active in the top-level category Tech & Innovation, the odds of mainly funding entrepreneurs who are active in Tech & Innovation increase by a factor of 8.76 compared to investors who are active in the omitted category Community Projects. Thus, the results of this model also support H4, as investors tend to invest in entrepreneurs with the same occupation when considering the top-level category Tech & Innovation.

Modell 4c: Funded Occupation—Community Projects

Also, the third model (see figure 11) supports hypothesis 4. Considering the funded occupation Community Projects as the dependent variable, the interaction with both expressions of the categorical independent variable for the backer occupation included in this model (i.e., Creative Works and Tech & Innovation) is negative and significant. Precisely, the interaction with the backer occupation Creative Works is negative and significant ($b = -3.98$, $p < .001$). Thus, for investors who are active in Creative Works, the odds of mainly funding entrepreneurs in the top-level category Community Projects change by a factor of 0.02, indicating that the probability significantly decreases compared to investors active in the omitted category Community Projects. Similar results are found for the interaction with the backer occupation Tech & Innovation, which is also negative and significant ($b = -2.75$, $p < .001$). This suggests that for investors whose occupation is Tech & Innovation, compared to investors whose occupation is Community Projects, the odds of mainly funding entrepreneurs in the top-level category Community Projects change by a factor of 0.06, indicating a decreasing probability. Since the probability decreases significantly in both cases, it can be conversely concluded that investors active in the omitted category Community Projects exhibit a bias to invest in entrepreneurs active in the same top-level category. Therefore, the results of this model also support hypothesis 4, as investors tend to support those entrepreneurs with whom they share the same occupation with, also in the case of the top-level category Community Projects.

Variables	b	Std Err	p
Independent Variable			
Backer race (White)	-0.31	0.35	0.38
Backer race (Black)	-0.33	0.51	0.51
Backer race (Asian)	-0.66	0.47	0.16
Backer race (Indian)	-0.92	0.83	0.27
Backer race (Latino Hispanic)	-0.88	0.58	0.13
Control Variables			
Backer sex	0.37	0.23	0.11
Backer age	0.02	0.02	0.28
Main category funded		Included	
Number of own projects	-0.08	0.11	0.48
Number of comments	0.00	0.00	0.56
Number of contributions	0.01*	0.00	0.02
Number of FBF	0.00	0.00	0.69
Facebook connected	-0.08	0.25	0.76
Twitter connected	0.34	0.26	0.20
Profile verified	-0.10	0.20	0.60
Profile word count	0.00	0.00	0.76

Notes. N = 1,246; ***$p < .001$; **$p < .01$; *$p < .05$; $^{\dagger} p < .1$

Figure 8: Logit model 3f for the dimension race (Race deviance—Middle Eastern)

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer occupation (Creative Works)	3.42***	0.28	0.00
Backer occupation (Tech & Innovation)	-0.07	0.32	0.83
Control Variables			
Backer sex	0.20	0.26	0.43
Backer race		Included	
Backer age	0.02	0.02	0.32
Number of own projects	0.08	0.13	0.55
Number of comments	-0.04*	0.02	0.02
Number of contributions	0.01	0.01	0.52
Number of FBF	0.00*	0.00	0.04
Facebook connected	0.13	0.24	0.58
Twitter connected	0.54*	0.27	0.05
Profile verified	-0.40†	0.22	0.06
Profile word count	0.00†	0.00	0.08

Notes. N = 830; ***$p < .001$; **$p < .01$; *$p < .05$; $^{\dagger} p < .1$

Figure 9: Logit model 4a for the dimension occupation (Creative Works)

Source: Own illustration

Dimension Location (H5)

In hypothesis 5, I predicted that contributions to entrepreneurs with a similar location as the investor occur more frequently than we would expect if investors chose projects for contribution at random. Thus, I predicted that the average distance between a backer and the entrepreneurs

Variables	b	Std Err	p
Independent Variable			
Backer occupation (Creative Works)	-1.15***	0.28	0.00
Backer occupation (Tech & Innovation)	2.17***	0.30	0.00
Control Variables			
Backer sex	-0.21	0.27	0.43
Backer race		Included	
Backer age	-0.01	0.02	0.66
Number of own projects	-0.09	0.13	0.50
Number of comments	0.04*	0.01	0.01
Number of contributions	-0.01	0.01	0.68
Number of FBF	0.00**	0.00	0.01
Facebook connected	-0.34	0.25	0.19
Twitter connected	-0.16	0.28	0.56
Profile verified	0.58*	0.23	0.01
Profile word count	0.00	0.00	0.36

*Notes. N = 830; ***p < .001; **p < .01; *p < .05; † p < .1*

Figure 10: Logit model 4b for the dimension occupation (Tech & Innovation)

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer occupation (Creative Works)	-3.98***	0.36	0.00
Backer occupation (Tech & Innovation)	-2.75***	0.38	0.00
Control Variables			
Backer sex	0.01	0.38	0.98
Backer race		Included	
Backer age	-0.03	0.03	0.30
Number of own projects	0.00	0.21	0.99
Number of comments	-0.01	0.02	0.74
Number of contributions	0.00	0.02	0.82
Number of FBF	0.00	0.00	0.33
Facebook connected	0.29	0.33	0.38
Twitter connected	-0.70†	0.38	0.07
Profile verified	-0.22	0.30	0.46
Profile word count	0.00	0.00	0.23

*Notes. N = 830; ***p < .001; **p < .01; *p < .05; † p < .1*

Figure 11: Logit model 4c for the dimension occupation (Community Projects)

Source: Own illustration

in the portfolio (*Actual average distance*) is significantly less than the average distance for the backer's expected portfolio (*Expected average distance*). That is, investors make homophilous choices with regard to the sociodemographic dimension location.

The result of the one-sided t-test fully supports hypothesis 5, as the test is highly significant ($p < .001$). Thus, I can conclude that the actual average distance between a backer and the entrepreneurs funded is significantly less than the expected average distance of the portfolio. Therefore, my results suggest that investors do not randomly choose entrepreneurs for contribution but instead exhibit a bias to invest in entrepreneurs to whom they are geographically close.

4.2. Robustness Test

As already mentioned, the sex of the individuals (i.e., of both backers and entrepreneurs) was additionally determined during data collection and processing using the genderize.io tool. Thus, I ran the same model as in chapter 4.1, but with the data from genderize.io, to test the robustness of hypothesis 2 and to validate whether the results are reproducible based on the alternative data.

Dimension Sex (H2)

In hypothesis 2, I predicted that contributions to entrepreneurs of the same sex as the investor occur more frequently than we would expect if investors chose projects for contribution at random. Thus, I predicted that backers make homophilous choices with respect to the sociodemographic characteristic sex.

Figure 12 presents the result from the logit regression model, which I ran as the robustness test to measure homophily in the dimension sex. These results also support hypothesis 2, as the interaction of the backer-sex with the corresponding sex deviance is positive and significant. Particularly for the model in figure 12, the interaction of the female backer-sex with the female sex deviance—indicating whether an investor supports more female entrepreneurs than expected—is positive and significant ($b = 0.88$, $p < .001$). This result suggests that for female investors compared to male investors, the odds to fund more female entrepreneurs than expected increase by a factor of 2.40. As already mentioned in chapter 4.1, it can therefore also be concluded that male investors support more male entrepreneurs than chance would suggest. Thus, these results reproduce the findings of the corresponding test of hypothesis 2 in chapter 4.1, implying that both female investors and male investors tend to support entrepreneurs of the same sex, indicating that backers make homophilous choices with regard to the dimension sex.

5. Discussion

In the past, numerous studies across various disciplines have shown the importance of the construct of homophily. However, most studies only considered one dimension of homophily, such as often the sociodemographic characteristic sex. Thus, previous studies neglected the complexity arising from the multidimensionality of this theoretical construct. Taking the first step to do more justice to the complexity by individually analyzing the five sociodemographic dimensions *age*, *sex*, *race*, *occupation*, and *location* based on the same dataset, I make several contributions.

First, I contribute to the broad literature on homophily. While my findings generally underline the importance of the theoretical construct, as various studies have demonstrated in the past, I add to the literature by showing that homophily is not as clear-cut considering more complex characteristics. That is the dimension race in the specific context of my study. Thus, I extend the existing literature by showing that bias depends on the particular sociodemographic characteristic of an individual under consideration. Thereby, for four of the five dimensions of homophily considered in my study, the results reproduce the findings of previous studies, indicating that homophily plays a vital role in individual choices. Therefore, these results will not be interpreted further here. It suffices to mention that I showed that individuals exhibit significant homophilous behavior in the dimensions age, sex, occupation, and location. That is, backers on the crowdfunding platform Indiegogo make homophilous choices in that they tend to invest in entrepreneurs who have a similar age and location as themselves, as well as in those entrepreneurs with whom they share the same sex and occupation.

More interesting are the results of my analyses in the sociodemographic dimension race, which will be interpreted in more detail in the following. In this context, I found mixed results, with significant differences depending on the race considered. In short, my findings suggest homophilous choices from individuals that are black or Indian. However, the same is not valid for white, Asian, Latino Hispanic, or Middle Eastern investors.

Specifically, my findings suggest that black and Indian entrepreneurs are discriminated against by all but their racial group, implying that on the other hand, backers who are black or Indian significantly tend to support more entrepreneurs of their race than chance would suggest, indicating homophilous choices. Furthermore, my findings provide more nuanced results for the white, Asian, Latino Hispanic, and Middle Eastern race. While in general, homophily seems to be relevant here as well, I do not find that all but their racial group significantly discriminate against those groups of entrepreneurs. Hence, I cannot conversely conclude that individuals from those racial groups exhibit significant homophilous behavior. For example, my findings for the white race suggest that Middle Eastern backers do not significantly discriminate against white entrepreneurs (compare model 3c). Similarly, my results show that Latino Hispanic and Indian investors do not significantly discriminate against Asian entrepreneurs (compare model 3d). Furthermore, Indian investors also do not significantly discriminate against Latino Hispanic entrepreneurs (compare model 3e). Finally, my findings suggest that Middle Eastern entrepreneurs are not significantly discriminated against by investors from any racial group (see model 3f). Therefore, I cannot conclude that, in general, white, Asian, Latino Hispanic, and Middle Eastern investors tend to make homophilous choices by funding more entrepreneurs of their race than expected. The above interpretation indicates that my analyses on racial homophily show different results depending on which of the six races is considered. In light of these findings, it could be interpreted as a result of cultural distance, which states that some cultures are more similar than other cultures (Shenkar, 2001), possibly explaining why certain races are discriminated against by some races but not by others.

Second, I contribute to the crowdfunding literature by

Variables	b	Std Err	p
Independent Variable			
Backer sex (Female)	0.88***	0.11	0.00
Control Variables			
Backer age	0.02*	0.01	0.01
Backer race		Included	
Main category funded		Included	
Number of own projects	0.16**	0.06	0.01
Number of comments	0.00	0.00	0.35
Number of contributions	0.01*	0.00	0.03
Number of FBF	0.00	0.00	0.47
Facebook connected	-0.14	0.13	0.30
Twitter connected	0.18	0.14	0.21
Profile verified	-0.15	0.11	0.15
Profile word count	0.01**	0.00	0.00

*Notes. N = 1,905; ***p < .001; **p < .01; *p < .05; † p < .1*

Figure 12: Logit model for the dimension sex—robustness test

Source: Own illustration

putting the investors into the spotlight of research. While previous studies in the crowdfunding context have mainly focused on the characteristics of the campaigns and the entrepreneurs as the most critical elements for success, past studies have neglected the third influential element: the role of the backers in crowdfunding. This gap is furthermore reflected by the scholars' call to also focus on the backers as a success factor in crowdfunding (Bretschneider & Leimeister, 2017; Kuppuswamy & Bayus, 2018). My study is one of the first to bring the investor into the spotlight of crowdfunding research, thus illuminating the component of crowdfunding success that has often been neglected. At the same time, my findings regarding the choice behavior of the backers also offer some valuable insights for crowdfunding entrepreneurs. As an example, in the dimension race, I demonstrate that, on the one hand, from the backers' perspective, only black and Indian backers make homophilous choices. On the other hand, from the entrepreneurial point of view, black and Indian entrepreneurs might be disadvantaged. While my findings do not permit to conclude that black and Indian entrepreneurs, in general, have fundamentally lower chances of crowdfunding success, I can undoubtedly say that investors of all other races discriminate against them. These findings may explain the emergence of crowdfunding platforms specifically for racial minorities, such as the platform FundBLACKFounders[7]. Furthermore, my findings suggest that, on average, younger entrepreneurs might be in a favorable position to achieve their funding goals. Since my results show that backers exhibit homophily in terms of age, and the average backer is about 31 years old (compare chapter 3.4), it could be interpreted that younger entrepreneurs have higher chances of success than entrepreneurs who are at the upper end of the age range. That is because there are simply more young backers in the crowd who tend to back entrepreneurs of similar age.

Third, I empirically contribute by obtaining the data for my analyses from Indiegogo, the second-largest crowdfunding platform. While most of the studies in crowdfunding use data from Kickstarter (exemplary see Greenberg and Mollick (2017); Kuppuswamy and Bayus (2018)), which is the largest crowdfunding platform, my study is one of the few using data from Indiegogo. Thus, my study offers new insights regarding investors' choices with respect to the backer-to-entrepreneur dyad based on a unique dataset gathered from Indiegogo.

The empirical contribution mentioned above is closely related to my fourth and final contribution. By replicating results found on other crowdfunding platforms, I contribute to the literature by providing an additional perspective on whether results found on one particular platform can be generalized across different crowdfunding platforms. That is because scholars have argued that crowdfunding platforms differ, and therefore, "patterns observed on one platform cannot be assumed to generalize to other platforms" (Dushnitsky & Fitza, 2018, p. 1). In this context, with my analyses based on data from Indiegogo, I reproduced previous results found on other platforms for resource acquisition online. For example, I was able to replicate the results of a study in the context of the lending-based microfinance platform Kiva. These findings suggest that lenders tend to fund fundraisers of the same sex and projects within an industry

[7]FundBLACKFounders is a reward based crowdfunding platform for black entrepreneurs founded to overcome the lack of funding allocated to black entrepreneurs (FundBLACKFounders, 2021).

similar to the one the lenders are active in (Riggins & Weber, 2017). Likewise, my results also suggest that investors tend to support entrepreneurs of the same sex and the same occupation. Just to mention a second example, this also holds for a study based on data from Kickstarter. Thereby, scholars were able to show that women support each other in crowdfunding (Greenberg & Mollick, 2017). This finding is also reflected in my study since my models suggest that investors tend to support entrepreneurs of the same sex, thus replicating these results by showing that female investors tend to fund female entrepreneurs.

6. Conclusion

This study sought out to analyze the effects of homophily. While I can, in principle, prove that investors on Indiegogo make homophilous choices, this does not apply to all dimensions of homophily tested for in my study. While investors primarily exhibit homophilous behavior with respect to the sociodemographic characteristics age, sex, occupation, and location, the answer is not as clear-cut in the more complex dimension race. Instead, significant differences are observable, depending on the race considered, where my results suggest that only black and Indian investors make significant homophilous choices. Despite offering valuable insights on how homophily influences the choice of backers in crowdfunding, this work nevertheless has some limitations that offer potential paths for future research. First, I have tried my best to reflect the mechanism underlying the backers' decision-making behavior. However, I cannot look in the investors' minds. Thus, I cannot judge whether the observed behavior was random or based on a conscious decision. Therefore, an exciting possibility for future research is a qualitative study in which detailed interviews are conducted with crowdfunding backers to understand better why they behave the way they do.

Second, there may be a slight bias in my samples, as only those individuals who uploaded a profile picture including their face were considered—which was caused by the fact that the characteristics of the individuals, such as their age and race, were algorithmically inferred from their profile pictures. All those individuals who did not upload a picture including their face were not taken into account in my study. This could cause a slight bias of reality, as scholars have shown that the face of the entrepreneur represents an important factor in the decision-making process of backers, as projects from entrepreneurs with higher facial trustworthiness have significantly more investors and a higher chance of success (Duan, Hsieh, Wang, & Wang, 2020). Future research might therefore, if possible, use alternative means to determine characteristics of individuals, such as the tools genderize.io, nationalize.io[8], and agify.io[9], to potentially have less biased samples, including not only those individuals where a face can be recognized from their profile pictures. Finally, another limitation arises from using the Python framework DeepFace to extract the characteristics of the individuals from their profile pictures. While technological progress nowadays allows a comparatively good estimation of the attributes of individuals based on their faces, these algorithms are not yet perfectly mature, whereby especially the determination of the age based on a face is still challenging.

Besides the topics mentioned above, my study creates additional exciting areas for future research. First, an increasing number of scholars in the field of homophily recently highlighted the need for studies on homophily that consider the multidimensionality of the concept (Block & Grund, 2014; Nahon & Hemsley, 2014). While my research provides a valuable first step in this direction by analyzing the influence of five different sociodemographic characteristics on homophily based on the same dataset gathered from the crowdfunding platform Indiegogo, further research is required to properly do justice to the complexity arising from the multidimensionality of this construct. In this context, a potentially fruitful path for future research is to apply a holistic view on homophily where different dimensions are not only studied individually but rather the interaction of the various dimensions is investigated.

Second, a more detailed analysis regarding the influence of the dimension race on homophily offers another exciting possibility for future research. That is because the dimension race is simply quite complex, with many possible combinations in a dyad. Since my findings suggest significant differences regarding racial homophily depending on the race considered, a detailed study of homophily considering the characteristic race is necessary to properly investigate which racial groups support each other.

Third, a more in-depth analysis regarding homophily in the dimension occupation is an attractive opportunity for future research. While I analyzed this dimension using the three top-level categories available on Indiegogo, future work could replicate the analyses using the more granular set of approximately 30 different categories (see Appendix 1). It would be interesting to determine whether the homophily hypothesis still holds in this case or whether, due to a higher level of complexity resulting from more possible variations, no strict answer can be given, just as I found for the dimension race.

Finally, replicating my results with data from other crowdfunding platforms is an interesting possibility for future research. On the one hand, this would further help to answer the question of platform generalizability (compare chapter 5). On the other hand, finding the same results on other crowdfunding platforms would additionally underline the robustness of the theoretical construct of homophily in the context of online venture funding. In conclusion, it should be said that my work is only a small step to advance research on homophily, especially in the context of crowdfunding, where I put the backers in the spotlight of research.

[8] Nationalize.io is a tool that predicts the nationality of a person given their name (Nationalize.io, 2021)

[9] Agify.io is a tool to predict the age of a person given their name (Agify.io, 2021)

I hope that in the future, scholars will follow this path to further advance research.

References

Abrams, D., & Hogg, M. A. (2006). *Social identifications: A social psychology of intergroup relations and group processes*. Routledge.

Agify.io. (2021). *Overview.* https://agify.io/#overview. (Accessed: 2021-07-18)

Ahuja, G., Polidoro Jr, F., & Mitchell, W. (2009). Structural homophily or social asymmetry? the formation of alliances by poorly embedded firms. *Strategic Management Journal*, *30*(9), 941–958.

Alba, R. D., & Golden, R. M. (1986). Patterns of ethnic marriage in the United States. *Social Forces*, *65*(1), 202–223.

Allison, T. H., Davis, B. C., Webb, J. W., & Short, J. C. (2017). Persuasion in crowdfunding: An elaboration likelihood model of crowdfunding performance. *Journal of Business Venturing*, *32*(6), 707-725.

Anglin, A. H., Wolfe, M. T., Short, J. C., McKenny, A. F., & Pidduck, R. J. (2018). Narcissistic rhetoric and crowdfunding performance: A social role theory perspective. *Journal of Business Venturing*, *33*(6), 780–812.

Aristotle. (1934). Rhetoric. Nichomachean ethics. In *Aristotle in 23 Volumes.* Harvard University Press.

Avery, D. R., McKay, P. F., & Wilson, D. C. (2007). Engaging the aging workforce: The relationship between perceived age similarity, satisfaction with coworkers, and employee engagement. *Journal of Applied Psychology*, *92*(6), 1542–1556.

Becker-Blease, J. R., & Sohl, J. E. (2007). Do women-owned businesses have equal access to angel capital? *Journal of Business Venturing*, *22*(4), 503–521.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. *Journal of Business Venturing*, *29*(5), 1–25.

Belleflamme, P., Omrani, N., & Peitz, M. (2015). The economics of crowdfunding platforms. *Information Economics and Policy*, *33*, 11–28.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure* (Vol. 7). Free Press.

Block, P., & Grund, T. (2014). Multidimensional homophily in friendship networks. *Network Science*, *2*(2), 189–212.

Bott, H. (1928). Observation of play activities in a nursery school. *Genetic Psychology Monographs*, *4*(1), 44–88.

Bourdieu, P. (1986). *The forms of capital*. Greenwood.

Bretschneider, U., & Leimeister, J. M. (2017). Not just an ego-trip: Exploring backers' motivation for funding in incentive-based crowdfunding. *The Journal of Strategic Information Systems*, *26*(4), 246–260.

Brigham, J. C., & Barkowitz, P. (1978). Do "They all look alike?" The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Social Psychology*, *8*(4), 306–318.

Byrne, D. E. (1971). *The attraction paradigm* (Vol. 462). Academic Press.

Carland, J. W., Hoy, F., & Carland, J. A. C. (1988). "Who is an entrepreneur?" is a question worth asking. *American Journal of Small Business*, *12*(4), 33–39.

Colombo, M. G., Franzoni, C., & Rossi–Lamastra, C. (2015). Internal social capital and the attraction of early contributions in crowdfunding. *Entrepreneurship Theory and Practice*, *39*(1), 75–100.

Cornelius, P. B., & Gokpinar, B. (2020). The role of customer investor involvement in crowdfunding success. *Management Science*, *66*(1), 452–472.

Cosh, A., Cumming, D., & Hughes, A. (2009). Outside enterpreneurial capital. *The Economic Journal*, *119*(540), 1494–1533.

Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, *54*(6), 2045–2073.

Coval, J. D., & Moskowitz, T. J. (2001). The geography of investment: Informed trading and asset prices. *Journal of Political Economy*, *109*(4), 811–841.

Cumming, D., & Dai, N. (2010). Local bias in venture capital investments. *Journal of Empirical Finance*, *17*(3), 362–380.

Dahlin, L., Rhue, L., & Clark, J. (2019). *Crowdfunding community formation: Fundraiser race and gender homophily*.

Davis, B. C., Hmieleski, K. M., Webb, J. W., & Coombs, J. E. (2017). Funders' positive affective reactions to entrepreneurs' crowdfunding pitches: The influence of perceived product creativity and entrepreneurial passion. *Journal of Business Venturing*, *32*(1), 90–106.

Deutsch, M. (1949). An experimental study of the effects of co-operation and competition upon group process. *Human Relations*, *2*(3), 199–231.

Duan, Y., Hsieh, T. S., Wang, R. R., & Wang, Z. (2020). Entrepreneurs' facial trustworthiness, gender, and crowdfunding success. *Journal of Corporate Finance*, *64*, 1–24.

Dushnitsky, G., & Fitza, M. (2018). Are we missing the platforms for the crowd? comparing investment drivers across multiple crowdfunding platforms. *Journal of Business Venturing Insights*, *10*, 1–10.

Feld, S. L. (1982). Social structural determinants of similarity among associates. *American Sociological Review*, *47*(6), 797–801.

FundBLACKFounders. (2021). *About.* https://www.fundblackfounders.com/about. (Accessed: 2021-07-13)

Genderize.io. (2021). *Overview.* https://genderize.io/#overview. (Accessed: 2021-04-20)

Giudici, G., Guerini, M., & Rossi-Lamastra, C. (2018). Reward-based crowdfunding of entrepreneurial projects: The effect of local altruism and localized social capital on proponents' success. *Small Business Economics*, *50*(2), 307–324.

Goldberg, C. B. (2005). Relational demography and similarity-attraction in interview assessments and subsequent offer decisions: Are we missing something? *Group & Organization Management*, *30*(6), 597–624.

Goldman, M., Stockbauer, J. W., & McAuliffe, T. G. (1977). Intergroup and intragroup competition and cooperation. *Journal of Experimental Social Psychology*, *13*(1), 81–88.

Gompers, P., Kovner, A., & Lerner, J. (2009). Specialization and success: Evidence from venture capital. *Journal of Economics & Management Strategy*, *18*(3), 817–844.

Greenberg, J., & Mollick, E. (2015). Leaning in or leaning on? gender, homophily, and activism in crowdfunding. *Academy of Management Proceedings*, *62*(2), 341–374.

Greenberg, J., & Mollick, E. (2017). Activist choice homophily and the crowdfunding of female founders. *Administrative Science Quarterly*, *62*(2), 341–374.

Greene, P. G., Brush, C. G., Hart, M. M., & Saparito, P. (2001). Patterns of venture capital funding: Is gender a factor? *Venture Capital: An International Journal of Entrepreneurial Finance*, *3*(1), 63-83.

Groza, M. P., Groza, M. D., & Barral, L. M. (2020). Women backing women: The role of crowdfunding in empowering female consumer-investors and entrepreneurs. *Journal of Business Research*, *117*, 432–442.

Guo, L., Guo, D., Wang, W., Wang, H., & Wu, Y. J. (2018). Distance diffusion of home bias for crowdfunding campaigns between categories: Insights from data analytics. *Sustainability*, *10*(4), 1251.

Hallinan, M. T., & Williams, R. A. (1989). Interracial friendship choices in secondary schools. *American Sociological Review*, *54*, 67–78.

Harrison, R. T., & Mason, C. M. (2007). Does gender matter? women business angels and the supply of entrepreneurial finance. *Entrepreneurship Theory and Practice*, *31*(3), 445–472.

Hegde, D., & Tumlinson, J. (2014). Does social proximity enhance business partnerships? Theory and evidence from ethnicity's role in US venture capital. *Management Science*, *60*(9), 2355–2380.

Hortaçsu, A., Martínez-Jerez, F., & Douglas, J. (2009). The geography of trade in online transactions: Evidence from eBay and mercadolibre. *American Economic Journal: Microeconomics*, *1*(1), 53–74.

Huberman, G. (2001). Familiarity breeds investment. *The Review of Financial Studies*, *14*(3), 659–680.

Huston, T. L., & Levinger, G. (1978). Interpersonal attraction and relationships. *Annual Review of Psychology*, *29*(1), 115–156.

Ibarra, H. (1995). Race, opportunity, and diversity of social circles in managerial networks. *Academy of Management Journal*, *38*(3), 673–703.

Indiegogo. (2021a). *About us.* https://www.indiegogo.com/about/our-story. (Accessed: 2021-03-24)

Indiegogo. (2021b). *Choose launch date & deadline.* https://support.indiegogo.com/hc/en-us/articles/205150367. (Accessed: 2021-03-25)

Indiegogo. (2021c). *Explore.* https://www.indiegogo.com/explore/all?project_type=campaign&project_timing=all&sort=trending. (Accessed: 2021-03-25)

Indiegogo. (2021d). *Pricing & fees.* https://learn.indiegogo.com/pricing-and-fees/. (Accessed: 2021-03-25)

Ingram, P., & Morris, M. W. (2007). Do people mix at mixers? structure, homophily, and the "life of the party". *Administrative Science Quarterly*, *52*(4), 558–585.
Jenq, C., Pan, J., & Theseira, W. (2015). Beauty, weight, and skin color in charitable giving. *Journal of Economic Behavior & Organization*, *119*, 234–253.
Kalmijn, M. (1998). Intermarriage and homogamy: Causes, patterns, trends. *Annual Review of Sociology*, *24*(1), 395–421.
Kalmijn, M., & Vermunt, J. K. (2005). Multidimensional homogeneity of social networks: A multilevel analysis of the role of age and marital status. *Social Networks*, *29*, 25–43.
Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology*, *84*(2), 427–436.
Kleinbaum, A. M., Stuart, T. E., & Tushman, M. L. (2013). Discretion within constraint: Homophily and structure in a formal organization. *Organization Science*, *24*(5), 1316–1336.
Kortum, S., & Lerner, J. (2000). Assessing the contribution of venture capital. *The RAND Journal of Economics*, *31*(4), 674–692.
Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology*, *115*(2), 405–450.
Kuppuswamy, V., & Bayus, B. L. (2018). Crowdfunding creative ideas: The dynamics of project backers. In *The economics of crowdfunding.* Cumming, D. and Hornuf.
Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. In *Freedom and control in modern society.* Berger, M. and Abel, T. and Page, C. H.
Lin, M., & Viswanathan, S. (2016). Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science*, *62*(5), 1393–1414.
Lincoln, J. R., & Miller, J. (1979). Work and friendship ties in organizations: A comparative analysis of relation networks. *Administrative Science Quarterly*, *24*(2), 181–199.
Loomis, C. P. (1946). Political and occupational cleavages in a Hanoverian village, Germany: A sociometric study. *Sociometry*, *9*(4), 316–333.
Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*(4), 330–334.
Massolution. (2015). *The crowdfunding industry report, 2015 CF.* https://www.smv.gob.pe/Biblioteca/temp/catalogacion/C8789.pdf.
McPherson, J. M. (1982). Hypernetwork sampling: Duality and differentiation among voluntary organizations. *Social Networks*, *3*(4), 225–249.
McPherson, J. M. (1983). An ecology of affiliation. *American Sociological Review*, *48*(4), 519–532.
McPherson, J. M., & Smith-Lovin, L. (1982). Women and weak ties: Differences by sex in the size of voluntary organizations. *American Journal of Sociology*, *87*(4), 883–904.
McPherson, J. M., & Smith-Lovin, L. (1986). Sex segregation in voluntary associations. *American Sociological Review*, *51*(1), 61–79.
McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American Sociological Review*, *52*(3), 370–379.
McPherson, J. M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.
Mehra, A., Kilduff, M., & Brass, D. J. (1998). At the margins: A distinctiveness approach to the social identity and social networks of underrepresented groups. *Academy of Management Journal*, *41*(4), 441–452.
Mobley, W. H. (1982). Supervisor and employee race and sex effects on performance appraisals: A field study of adverse impact and generalizability. *Academy of Management Journal*, *25*(3), 598–606.
Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, *29*(1), 1–16.
Mollick, E., & Robb, A. (2016). Democratizing innovation and capital access: The role of crowdfunding. *California Management Review*, *58*(2), 72–87.
Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, *23*(2), 242–266.
Nahon, K., & Hemsley, J. (2014). Homophily in the guise of cross-linking: Political blogs and content. *American Behavioral Scientist*, *58*(10), 1294–1313.
Nationalize.io. (2021). *Overview.* https://nationalize.io/#overview. (Accessed: 2021-07-18)
Oo, P. P., Allison, T. H., Sahaym, A., & Juasrikul, S. (2019). User entrepreneurs' multiple identities and crowdfunding performance: Effects through product innovativeness, perceived passion, and need similarity. *Journal of Business Venturing*, *34*(5), 1–49.
Plato. (1968). Laws [Book Section]. In *Plato in twelve volumes* (Vol. 11). Bury translator. Cambridge: Harvard University Press.
Powell, W. W., White, D. R., Koput, K. W., & Owen-Smith, J. (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology*, *110*(4), 1132–1205.
Python Package Index. (2021a). *Project description - DeepFace.* https://pypi.org/project/deepface/. (Accessed: 2021-04-20)
Python Package Index. (2021b). *Project description - Geopy.* https://pypi.org/project/geopy/. (Accessed: 2021-04-20)
Python Package Index. (2021c). *Project description - Haversine.* https://pypi.org/project/haversine/. (Accessed: 2021-04-20)
Richardson, H. M. (1940). Community of values as a factor in friendships of college and adult women. *The Journal of Social Psychology*, *11*(2), 303–312.
Riggins, F., & Weber, D. (2017). Information asymmetries and identification bias in P2P social microlending. *Information Technology for Development*, *23*(1), 107–126.
Scheaf, D. J., Davis, B. C., Webb, J. W., Coombs, J. E., Borns, J., & Holloway, G. (2018). Signals' flexibility and interaction with visual cues: Insights from crowdfunding. *Journal of Business Venturing*, *33*(6), 720–741.
Scrapy. (2021). *Documentation.* https://docs.scrapy.org/en/latest/. (Accessed: 2021-04-23)
Shenkar, O. (2001). Cultural distance revisited: Towards a more rigorous conceptualization and measurement of cultural differences. *Journal of International Business Studies*, *32*(3), 519–535.
Shrum, W., Cheek Jr, N. H., & MacD, S. (1988). Friendship in school: Gender and racial homophily. *Sociology of Education*, *61*(4), 227–239.
Smith, J. A., McPherson, M., & Smith-Lovin, L. (2014). Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. *American Sociological Review*, *79*(3), 432–456.
Sohl, J. E. (2015). *The angel investor market in 2014: A market correction in deal size*.
Soublière, J. F., & Gehman, J. (2020). The legitimacy threshold revisited: How prior successes and failures spill over to other endeavors on Kickstarter. *Academy of Management Journal*, *63*(2), 472–502.
Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In *Advances in experimental social psychology* (pp. 261–302). Berkowitz, L.
Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of Social Issues*, *41*(3), 157–175.
Stuart, T. E., & Sorenson, O. (2007). Strategic networks and entrepreneurial ventures. *Strategic Entrepreneurship Journal*, *1*(3-4), 211–227.
Taeuscher, K., Bouncken, R. B., & Pesch, R. (2020). Gaining legitimacy by being different: Optimal distinctiveness in crowdfunding platforms. *Academy of Management Journal*, *64*(1), 149–179.
Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information*, *13*(2), 65–93.
Tan, Y. H., & Reddy, S. K. (2021). Crowdfunding digital platforms: Backer networks and their impact on project outcomes. *Social Networks*, *64*, 158–172.
Thelwall, M. (2009). Homophily in myspace. *Journal of the American Society for Information Science and Technology*, *60*(2), 219–231.
Tsui, A. S., & Gutek, B. A. (1999). *Demographic differences in organizations: Current research and future directions*. Lexington Books.
Tsui, A. S., & O'reilly, C. A. (1989). Beyond simple demographic effects: The importance of relational demography in superior-subordinate dyads. *Academy of Management Journal*, *32*(2), 402–423.
Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory* [Book]. Basil Blackwell.
Turner, J. C., & Oakes, P. J. (1986). The significance of the social identity

concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology*, *25*(3), 237–252.

Verbrugge, L. M. (1977). The structure of adult friendship choices. *Social Forces*, *56*(2), 576–597.

Verbrugge, L. M., Gruber-Baldini, A. L., & Fozard, J. L. (1996). Age differences and age changes in activities: Baltimore longitudinal study of aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *51*(1), S30–S41.

Vismara, S., Benaroio, D., & Carne, F. (2017). Gender in entrepreneurial finance: Matching investors and entrepreneurs in equity crowdfunding. In *Gender and entrepreneurial activity* (pp. 271–288). Link, A. N. and Phillips, V. B.

Wellman, B. (1996). Are personal communities local? a dumptarian reconsideration. *Social Networks*, *18*(4), 347–354.

Wimmer, A., & Lewis, K. (2010). Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *American Journal of Sociology*, *116*(2), 583–642.

World Bank. (2013). *Crowdfunding's potential for the developing world.* https://www.infodev.org/infodev-files/wb_crowdfundingreport-v12.pdf.

Younkin, P., & Kuppuswamy, V. (2018). The colorblind crowd? founder race and performance in crowdfunding. *Management Science*, *64*(7), 3269–3287.