Mirrored Matching: Facial Similarity and the Allocation of Venture Capital

Emmanuel Yimfor

Columbia University
665 W 130th St, New York, New York 10027, U.S.A.

September 23, 2025

Abstract

Do subtle visual cues influence high-stakes economic decisions? Using venture capital as a laboratory, this paper shows that facial similarity between investors and entrepreneurs predicts positive funding decisions but negative investment outcomes. Analyzing early-stage deals from 2010-2020, we find that greater facial resemblance increases match probability by 3.2 percentage points even after controlling for same race, gender, and age, yet funded companies with similar-looking investor-founder pairs have 7 percent lower exit rates. However, when deal sourcing is externally curated, facial similarity effects disappear while demographic homophily persists, indicating facial resemblance primarily operates as an initial screening heuristic. These findings reveal a novel form of homophily that systematically shapes capital allocation, suggesting that interventions targeting deal sourcing may eliminate the negative influence of visual cues on investment decisions.

JEL: G24, J15, L26

Keywords: Venture Capital, Facial Similarity, Image Embeddings, Diversity, Bias, Investment Matching, Homophily, Social Networks, Machine Learning, Behavioral Bias

* I would like to thank Kelly Shue, Laura Velkamp, Matt Marx, and Heather Tookes for their comments on an earlier version of the draft. All errors are mine.
Email address: emmanuel.yimfor@columbia.edu (Emmanuel Yimfor)

Mirrored Matching: Facial Similarity and the Allocation of Venture Capital

Abstract

Do subtle visual cues influence high-stakes economic decisions? Using venture capital as a laboratory, this paper shows that facial similarity between investors and entrepreneurs predicts positive funding decisions but negative investment outcomes. Analyzing early-stage deals from 2010-2020, we find that greater facial resemblance increases match probability by 3.2 percentage points even after controlling for same race, gender, and age, yet funded companies with similar-looking investor-founder pairs have 7 percent lower exit rates. However, when deal sourcing is externally curated, facial similarity effects disappear while demographic homophily persists, indicating facial resemblance primarily operates as an initial screening heuristic. These findings reveal a novel form of homophily that systematically shapes capital allocation, suggesting that interventions targeting deal sourcing may eliminate the negative influence of visual cues on investment decisions.

JEL: G24, J15, L26, D85, G24

1. Introduction

In a world of incomplete information, how do decision makers navigate high-stakes decisions that shape economies, organizations, and individual trajectories? From medical diagnoses to investment allocations, hiring choices to partner selection, decision-makers face the persistent challenge of predicting distant, probabilistic outcomes based on limited data. In the face this uncertainty, humans sometimes instinctively resort to subtle perceptual signals that sometimes operate beneath conscious awareness (Tversky and Kahneman, 1974; Kahneman and Tversky, 1979; Gigerenzer and Gaissmaier, 2011). These hidden influences may systematically shape who receives capital and opportunities.

This paper tests the hypothesis that facial resemblance predicts matching outcomes, using the venture capital ecosystem as a laboratory where information asymmetry is extreme and matching decisions have substantial economic consequences. We will address three core questions. First, does facial similarity predict matching of founders and venture partners? Second, where in the multi-stage investment process (initial sourcing or final selection) is it most important? Third, does facial similarity predict better or worse investment outcomes? Understanding these distinctions has implications for capital allocation efficiency.

To test where facial similarity is most influential, we will contrast typical investment patterns with those observed in contexts where an intermediary largely shapes the initial consideration set. This approach helps determine if facial similarity primarily operates during an investor's own deal sourcing and initial screening phases or during the later stages of final selection. If facial similarity mainly affects final investment decisions, its impact should largely persist even when the initial opportunity pool is externally curated. Conversely, if it functions primarily as an early screening heuristic, its influence should diminish or disappear when such initial filtering is conducted externally.

Our analysis uses data from PitchBook from 2010 to 2020 and concentrates on early-stage venture capital deals, a stage where the high uncertainty inherent in early-stage investments may make heuristics, such as facial similarity, more influential in decision-making. For each deal, we identify the lead investment partner and the primary founder (typically the CEO) of the funded startup, collecting facial images for these individuals from public sources (Cook et al., 2024). To simulate the investor's decision process, we construct counterfactual matches. These matches represent plausible alternative startups that each investor likely considered but ultimately did not fund. Following established sourcing and screening patterns, we implement a three-

step procedure to identify these counterfactual opportunities (Gompers et al., 2020a).

First, following the literature, we match on fundamental deal characteristics: deal type (exact match), industry (exact match), company founding year (exact match), and geographic location (exact match at state level), reflecting how investors typically focus on specific investment theses (Gompers et al., 2016, 2020b; Garfinkel et al., 2021). Second, we restrict potential matches to deals occurring within a 180-day window of the actual deal, capturing the contemporaneous opportunity set an investor would evaluate during their decision-making process. Third, among these potential matches, we select the deal with the closest deal size to ensure economic similarity.

This process effectively creates a comparison portfolio for each investor, containing both the startup they actually funded and a highly similar startup they plausibly considered but did not fund. This paired approach yields 43,204 actual deals and an equal number of counterfactual matches, allowing us to test the hypothesis that facial similarity between the partner and founder predicts matching.

Using machine learning-based FaceNet embeddings from the 86,409 potential investor-founder pairs, we show that facial resemblance significantly increases the likelihood of matching, even after fixing demographic characteristics, network connections, and deal attributes. A one standard deviation increase in facial similarity raises match probability by 3.2 percentage points relative to the unconditional mean.

Facial similarity matters much more for non-white investors (3.4× larger), female investors (2.2× larger), and when people share networks (1.9× larger).[1] If underrepresented investors face higher information barriers evaluating underrepresented founders, they may rely more on facial similarity as a shortcut or as a proxy for shared traits. These findings imply that the observed reliance on facial similarity extends beyond simple demographic homophily, possibly reflecting a nuanced response to varying informational environments where investors might use cues related to appearance to navigate uncertainty.

Next, we test whether facial similarity predicts better or worse investment outcomes by examining the subsequent performance of funded companies. Facial similarity predicts worse investment outcomes. Startups with facially similar matches are 7% less likely to have successful exits. This suggests that facial similarity is not a useful signal of startup quality.

To understand the mechanisms through which facial similarity predicts matching,

[1] Shared network connections, which capture whether a founder and lead partner previously overlapped at the same company or university, significantly increase the likelihood of matching (Cook et al., 2024).

we turn to a context where the initial consideration set is external curated by an intermediary. When we examine matching between startups selected by Y combinator and lead partners, we find that facial similarity does not predict matching, even though demographic and network effects continue to predict matching. This pattern suggests that facial similarity mainly operates during the screening phase, not in final selection decisions. Our hypothesis is that Y Combinator's certification likely substitutes for facial heuristics, providing investors with stronger quality signals than appearance. This evidence suggests facial similarity functions primarily as an initial screening tool that helps investors narrow their consideration sets.

This paper makes two contributions to the study of decision-making and capital allocation. First, it introduces facial similarity as a new measure of homophily. Prior work focuses on categorical traits such as race, gender, or education (Bengtsson and Hsu, 2015; Garfinkel et al., 2021; Gompers et al., 2017; Cook et al., 2024). We show that facial similarity strongly predicts investor–founder matching even after controlling for observable demographic, geographic, and network characteristics. The association is especially pronounced among underrepresented investors, where information frictions may be greater. Second, by comparing standard VC deals to those sourced through Y Combinator, where founder selection is externally constrained, we show that different dimensions of homophily operate through distinct mechanisms. While facial similarity effects disappear in the Y Combinator context, race and gender effects persist. This divergence suggests that facial similarity functions primarily as a heuristic for narrowing consideration sets during initial screening, whereas demographic homophily along race and gender lines reflects deeper partner-founder preferences or network affinities that persist even when deals are externally sourced (Cook et al., 2024). Together, these findings identify facial similarity as a potentially informative marker of homophily that shapes early-stage capital allocation, while distinguishing its role as a screening mechanism from the more persistent influence of categorical demographic characteristics.

This research also contributes to the literature on appearance-based biases in economic decision-making, but with a fundamental distinction. Prior work examines how universal appearance traits affect outcomes. Brooks et al. (2014) find investors prefer ventures pitched by attractive men, Duarte et al. (2012) show trustworthy-looking borrowers receive better lending terms, and Huang et al. (2023) demonstrate that perceived personality traits from photographs influence investment decisions. These studies focus on how specific appearance attributes are universally valued or devalued. In contrast, we examine facial similarity between decision-makers and recipients, a relational measure rather than an absolute trait. Our findings suggest homophily

effects: investors are more likely to fund founders who look like them, with this effect varying across investor demographics. This distinction matters because it reveals a different mechanism of appearance-based bias, one driven by *relative* rather than universal preferences.

Moreover, facial similarity may capture deeper homophily beyond surface-level appearance. Extensive research shows facial features correlate with personality traits (Todorov and Porter, 2015), cognitive processing styles (Bowles et al., 2022), political orientations (Kosinski, 2021), and career-relevant characteristics (Guenzel et al., 2025). Thus, when investors select facially similar founders, they may be responding to subtle cues about shared psychological or cognitive traits rather than, or in addition to, visual resemblance itself. However, our outcome analysis reveals that whatever signals facial similarity captures, they lead to worse investment decisions. Founder-partner pairs matched on facial similarity are 7% less likely to achieve successful exits, indicating that these appearance-based cues, whether reflecting personality, cognitive styles, or other traits, constitute a costly bias rather than valuable information for investment selection.

The remainder of this paper proceeds as follows. Section 2 details the facial embedding techniques used to quantify similarity between investors and founders. Section 4 describes the data sources and sample construction, including the counterfactual matching approach. Section 3 outlines the empirical strategy, including the main specification and additional identification approaches. Section 5 presents the findings on facial similarity and investment matching, including heterogeneous effects across investor demographics. Section 6 concludes.

2. Methodology

This paper uses facial embeddings to quantify similarity between venture capital partners and startup founders. To measure facial similarity, we use FaceNet, a state-of-the-art deep learning facial recognition system developed by Google Research. The methodology involves several steps: image preprocessing, facial embedding extraction, and similarity score calculation.

2.1. Image Preprocessing

Before extracting facial embeddings, we preprocess the images for consistency. The process begins with face detection using *RetinaFace*, a robust face detection algorithm that identifies the precise location of faces within images. Once detected, each face is aligned based on facial landmarks (eyes, nose, mouth) to normalize pose variations

across images. The aligned faces are then cropped to focus solely on the facial region and resized to 160×160 pixels. This preprocessing standardizes the facial images by removing background elements and focusing only on facial features, correcting for variations in head tilt and orientation, normalizing for distance from the camera, and creating uniform input dimensions for the neural network model.

The importance of this standardization is illustrated in Figure 1, which presents processed images of three female celebrities. As shown in the figure, despite differences in the original photos (background, lighting, pose), the standardized images isolate facial features in a consistent format that enables direct comparisons.

[INSERT FIGURE 1 ABOUT HERE.]

2.2. Facial Embedding Extraction

After preprocessing, the standardized facial images are passed through the *FaceNet* neural network model, which maps each face to a 128-dimensional embedding vector. The resulting 128-dimensional embedding represents a face's distinctive characteristics in a compact numerical format. The embedding captures both broad facial structures and fine-grained features, including aspects related to facial geometry and proportions, eye, nose, and mouth shapes, skin texture patterns, facial symmetry, and subtle contours and distinctive features.

These embeddings enable quantitative comparisons between faces, providing a mathematical representation of facial features that goes beyond what might be immediately apparent.

2.3. Similarity Score Calculation

To quantify similarity between two faces, we calculate the cosine similarity between their respective 128-dimensional *FaceNet* embeddings. Cosine similarity yields a value between negative one and one, where one indicates perfect similarity and negative one indicates perfect dissimilarity. We further normalize the cosine similarity to range from zero (perfect dissimilarity) to one (perfect similarity). For each partner-founder (PF) pair, the facial similarity score is:

$$\text{similarity}(P, F) = \frac{P \cdot F}{||P|| \cdot ||F||} = \frac{\sum_{i=1}^{128} P_i F_i}{\sqrt{\sum_{i=1}^{128} P_i^2} \sqrt{\sum_{i=1}^{128} F_i^2}}, \tag{1}$$

Where P and F represent the 128-dimensional embeddings of the partner and founder, respectively.

As we see in Figure 1, the preprocessing ensures that the resulting 128-dimensional embeddings capture intrinsic facial geometry rather than incidental photographic variations. The similarity scores displayed between each pair, ranging from 0.6281 to 0.8403, demonstrate how this standardization allows FaceNet to detect nuanced differences in facial structure. Notably, the twins Mary-Kate and Ashley Olsen show the highest similarity (0.8403), while their younger sister Elisabeth exhibits lower but still substantial similarity to each (0.7389 with Mary-Kate, 0.6281 with Ashley). These within-gender, within-race comparisons validate that the embeddings capture similarities beyond the broad demographic categories.

3. Empirical Strategy

3.1. Validating the Facial Similarity Measure

To validate our measure of facial similarity and understand how much of it is explained by demographic factors, we first estimate:

$$FacialSimilarity_{ij} = \alpha + \beta_1 SameRace_{ij} + \beta_2 SameGender_{ij} + \beta_3 SimilarAge_{ij} + \delta_t + \epsilon_{ij}. \tag{2}$$

This decomposition allows us to verify that facial similarity captures meaningful variation beyond coarse demographic categories, which is essential for our identification strategy.

3.2. Main Estimation Approach

Our empirical strategy relies on comparing actual investment matches to counterfactual matches that share similar observable characteristics. The main specification for testing whether facial similarity predicts matching is:

$$Match_{ij} = \alpha + \beta FacialSimilarity_{ij} + \gamma X_{ij} + \delta_t + \theta_p + \epsilon_{ij}, \tag{3}$$

where $Match_{ij}$ is an indicator equal to 100 for actual matches and 0 for counterfactual matches between investor i and founder j, $FacialSimilarity_{ij}$ is the standardized facial similarity score, X_{ij} includes control variables such as same race, same gender, similar age, and shared network connections. The model incorporates year fixed effects (δ_t) to control for market-wide temporal trends, and partner fixed effects (θ_p) to account for time-invariant investor characteristics including risk preferences and investment style.

3.3. Heterogeneous Effects

To examine heterogeneous effects across investor demographics, we estimate:

$$Match_{ij} = \alpha + \beta FacialSimilarity_{ij} + \lambda FacialSimilarity_{ij} \times InvestorDemographic_i + \gamma X_{ij} + \delta_t + \theta_p + \epsilon_{ij}, \tag{4}$$

where $InvestorDemographic_i$ are binary indicators for investor characteristics such as being non-white or female.

3.4. Investment Outcomes

To test whether facial similarity serves as a proxy for unobserved founder quality, we examine the relationship between facial similarity and subsequent venture performance. If facial similarity captures meaningful signals about founder quality or investor-founder compatibility, we would expect it to positively predict investment outcomes. Conversely, if it reflects a bias that leads investors away from optimal decisions, we would expect negative performance effects. To that end, we estimate:

$$Outcome_{ij} = \alpha + \beta FacialSimilarity_{ij} + \gamma X_{ij} + \delta_t + \theta_p + \epsilon_{ij} \tag{5}$$

Where $Outcome_{ij}$ is an indicator for successful exit via IPO or acquisition where the acquisition price exceeds twice the cumulative funding raised. This exit measure follows Yimfor and Garfinkel (2023), who show that successful exits strongly correlate with other performance metrics including follow-on funding success, valuation growth, and long-term survival. By focusing on this comprehensive exit measure, we capture the ultimate success metric that subsumes these intermediate outcomes. The coefficient β indicates whether facial similarity predicts better or worse investment outcomes after controlling for observable characteristics, X_{ij}.

3.5. Y Combinator Setting

To understand whether facial similarity reflects network-based sourcing effects or final-stage evaluation preferences, we leverage Y Combinator's accelerator program. In traditional VC investing, investors both source deals (deciding which startups to consider) and make final investment decisions. Y Combinator breaks this process in two: they handle the initial screening through their competitive application process, while investors only engage in the final selection stage. This separation allows us to isolate whether facial similarity operates primarily as a sourcing heuristic or continues to influence final investment decisions.

For the sample of startups accepted into each YC batch, we estimate following regression:

$$Match_{ij,YC} = \alpha + \beta FacialSimilarity_{ij} + \gamma X_{ij} + \delta_t + \theta_p + \epsilon_{ij} \tag{6}$$

If facial similarity primarily operates through network-based sourcing channels, we would expect β to be near zero in this setting.

3.6. Limitations and Interpretation

We cannot claim our estimates are causal. Even with partner fixed effects and careful matching, unobserved factors might drive both facial similarity and investment decisions. Still, understanding how facial similarity correlates with investment outcomes sheds light on the capital allocation process. We will estimate robust correlations that hold after controlling for demographics, networks, and deal characteristics. Given the enormous economic consequences of venture capital decisions and the opacity of how investors winnow hundreds of potential deals down to the few they fund, these patterns shed light on a critical but poorly understood screening process (Gompers et al., 2020a).

4. Data

The primary data comes from PitchBook, a comprehensive database that tracks private capital markets including venture capital investments. We focus on early-stage deals (angel investments, seed rounds, and early-stage VC) where information asymmetry between investors and founders is particularly severe. At these early stages, startups typically lack extensive operating histories, validated business models, or meaningful financial metrics that characterize later-stage investments. This extreme uncertainty creates an environment where heuristics like facial similarity might shape investment decisions. Our sample covers deals occurring between 2010 and 2020, which represents a period of significant growth and evolution in the venture capital ecosystem, providing substantial variation in both investment patterns and outcomes to test our hypotheses.

For each deal, we identify the lead partner at the investing firm and the CEO or primary founder of the startup receiving funding. Next, we collect images for these individuals from public sources, primarily company websites and LinkedIn profiles. Following this approach from Cook et al. (2024), we found images for about 95 percent of partners and founders in the sample. Finally, we process these images following the methodology described in Section 2.

To derive demographic information, we follow the approach described in (Cook et al., 2024). We classify each founder or partner by race (White, Black, Asian, Hispanic), gender (Male, Female), and age (categorized as 25-35, 36-45, 46-55, 56-65, and 66-75 years). The deal characteristics data includes deal type (angel, seed, early-stage VC), deal size in millions of dollars, industry classification, company founding year, geographic location (state), and deal date.

To examine network connections between investors and founders, we collected LinkedIn profile data for both groups. This allowed us to identify shared educational institutions and overlapping work histories. Specifically, we determine whether a founder and investor attended the same university or worked at the same company prior to the investment deal. These network connections are crucial controls given the importance of warm introductions and referrals in venture capital (Gompers et al., 2020b).

For our analysis of how facial similarity operates in different contexts, we also constructed a sample of Y Combinator participants. We collected data on all companies that participated in Y Combinator from 2010 to 2023. We then matched these companies to PitchBook using company names and founding dates. This Y Combinator sample allows us to test whether facial similarity effects persist when an intermediary (Y Combinator) conducts initial screening.

For the Y Combinator analysis, we construct counterfactual matches differently to reflect the structured nature of YC's program. We identify all investments made by lead partners in YC companies within 180 days following each batch's demo day. For each lead partner who invested in a YC company, we create a consideration set consisting of all other companies from the same YC batch operating in the same industry sector. This approach reflects the reality that YC investors typically evaluate entire cohorts within their sector focus during demo days. For example, if a partner invested in a fintech company from the Winter 2015 batch, their counterfactual set includes all other fintech companies from that same batch. This process yields 15,128 potential partner-founder pairs, where actual investments are compared against all feasible alternatives within the same batch-industry cell.

While PitchBook provides data on realized investments, understanding how facial similarity influences matching requires comparing actual matches with potential matches that did not occur. To estimate whether facial similarity predicts investor-founder matching, we need to simulate the choice set that investors faced when making their investment decisions. Therefore, to construct a counterfactual sample that represents plausible alternative investments each investor could have made, we follow a three-step procedure.

First, we match on fundamental deal characteristics: deal type (exact match), industry (exact match), company founding year (exact match), and geographic location (exact match at state level). Second, we restrict potential matches to deals occurring within a 180-day window of the actual deal to ensure temporal comparability. Third, among these potential matches, we select the deal with the closest deal size to ensure economic similarity. This process yields 43,204 actual deals and an equal number of counterfactual matches.

5. Results

5.1. Summary Statistics

Table I presents summary statistics comparing actual investment matches with counterfactual matches. The table shows several interesting patterns. First, facial similarity scores are significantly higher between founders and investors in actual matches compared to counterfactual matches, with a mean standardized facial similarity (standardized to have a mean of zero and standard deviation of one) of 0.03 in actual deals versus -0.04 in counterfactual deals. This univariate evidence suggests that facial similarity may play a role in investment decisions.

[INSERT TABLE I ABOUT HERE.]

Second, demographic similarities are also more prevalent in actual matches. Same-race pairs appear in 65% of actual matches versus 60% of counterfactual matches. Similarly, same-gender pairs are more common in actual matches (84% versus 82%), as are similar-age pairs (34% versus 31%). Beyond demographics, shared network connections also show a strong homophily effect: 7% of actual matches involve partners and founders who previously overlapped at the same university or company, compared to only 3% in counterfactual matches. These patterns suggest homophily across multiple dimensions in venture capital matching. However, the key empirical question is whether facial similarity predicts matching beyond these coarse demographic categories. If investors simply prefer founders of the same race or gender, facial similarity might offer no additional explanatory power. Alternatively, facial similarity could capture more nuanced preferences within demographic groups, revealing a finer-grained dimension of homophily in investment decisions.

Third, the distributions of investor demographics are balanced across actual and counterfactual matches, with non-white investors comprising 25% and female investors 9% of both samples. This balance confirms the effectiveness of the matching procedure in creating comparable groups that differ primarily in whether a match occurred.

Finally, the matching procedure successfully balanced deal characteristics: both groups average $9.20 million deal sizes with identical distributions across deal types and industries. This balance strengthens the validity of subsequent analyses by ensuring that differences in matching outcomes are not driven by underlying differences in deal attributes.

5.2. Facial Similarity Within and Across Groups

Table I raised a crucial question: does facial similarity merely reflect coarse demographic preferences, or does it capture more nuanced variation within groups? Figure 2 provides initial evidence. The distribution of facial similarity scores by same-race versus different-race pairs shows that while same-race pairs have higher average similarity, there is substantial overlap between distributions. This within-race variation, where some same-race pairs are less similar than many different-race pairs, confirms that our measure captures meaningful facial differences beyond demographic categories. This variation is essential for our identification strategy, allowing us to test whether facial similarity predicts matching even after controlling for race, gender, and age.

[INSERT FIGURE 2 ABOUT HERE.]

5.3. Determinants of Facial Similarity

To quantify the within-group variation shown in Figure 2, Table II estimates how much of facial similarity is explained by demographic categories using equation 2. This multivariate approach allows us to decompose facial similarity into the portion driven by race, gender, and age versus the residual variation within these groups.

[INSERT TABLE II ABOUT HERE.]

The results show that shared demographic characteristics significantly predict higher facial similarity. Specifically, same-race pairs have facial similarity scores that are 0.228 standard deviations higher than different-race pairs (Column 1). Same-gender pairs have similarity scores 0.385 standard deviations higher than different-gender pairs (Column 2), and similar-age pairs have scores 0.138 standard deviations higher (Column 3).

When including all three demographic factors together (Column 4), each remains significant with only slight reductions in coefficient magnitudes. This suggests that race, gender, and age contribute independently to facial similarity. The adjusted R-squared of 0.039 in the full model indicates that while demographic factors meaningfully predict facial similarity, most of the variation in facial similarity exists within

demographic groups. This within-group variation is crucial for the subsequent analyses, where we test whether facial similarity predicts investment matching controlling for demographic similarities.

5.4. Facial Similarity and Investment Match Probability

Table III presents the main results from estimating equation 3, which tests how facial similarity predicts investment match probability. Column 1 shows that a one standard deviation increase in facial similarity is associated with a 2.443 percentage point increase in the probability of an investment match. This effect represents a substantial 4.9% increase relative to the baseline 50% probability in the matched sample.

[INSERT TABLE III ABOUT HERE.]

Columns 2 through 5 progressively add demographic controls to assess whether facial similarity has an independent effect beyond demographic homophily. After controlling for same race, same gender, similar age, and shared network connections, the coefficient on facial similarity remains economically and statistically significant at 1.576 (Column 5). This indicates that facial similarity predicts investment matches even among pairs that share the same demographic characteristics and network connections.

The control variables themselves reveal substantial homophily effects. Having the same race increases match probability by 7.414 percentage points, same gender by 5.079 percentage points, and similar age by 3.826 percentage points. Shared network connections have the largest effect, increasing match probability by 26.997 percentage points. Despite these strong homophily effects, facial similarity remains a significant predictor, suggesting it captures an additional dimension of similarity that predicts matching in the venture capital market.

Having established that facial similarity predicts investment matching beyond demographic homophily, we next test whether this effect varies across investor characteristics and whether it leads to better or worse investment outcomes.

5.5. Heterogeneous Effects Across Investor Demographics

Table IV presents results from estimating equation 4, which tests how the effect of facial similarity varies across investor demographics. Column 1 shows a strong interaction between facial similarity and investor race. While facial similarity has a

positive effect for white investors (coefficient of 0.844), this effect is substantially larger for non-white investors (additional effect of 2.825, making the total effect 3.4 times larger). Column 2 shows a similar pattern for investor gender, with the effect for female investors (additional effect of 1.935) approximately 2.3 times larger than for male investors.

[INSERT TABLE IV ABOUT HERE.]

Column 3 examines the interaction with shared network connections. The coefficient on the interaction term (1.950) indicates that facial similarity has a stronger effect on match probability when the investor and founder share network connections. Column 4 includes all interaction terms simultaneously, and the results remain qualitatively similar, suggesting that each interaction represents an independent effect.

Figure 3 visualizes these heterogeneous effects using binned scatter plots. Panel A shows that the relationship between facial similarity and match probability is steeper for non-white investors compared to white investors, especially at higher levels of facial similarity. Similarly, Panel B shows a steeper relationship for female investors compared to male investors. These visualizations reinforce the finding that facial similarity has stronger effects on investment decisions for investors from underrepresented groups.

[INSERT FIGURE 3 ABOUT HERE.]

This finding suggests that underrepresented investors may face higher information barriers when evaluating underrepresented founders. Therefore, they may rely more heavily on facial similarity as a heuristic for shared background, values, or communication styles.

5.6. Facial Similarity and Investment Outcomes

Having established that facial similarity predicts investment matching, we now examine whether this matching pattern reflects valuable information or costly bias. If facial similarity captures unobserved founder quality that our controls miss (such as cognitive ability, work ethic, or leadership potential) we would expect it to positively predict investment outcomes. Venture capitalists might reasonably infer that founders who resemble them share similar analytical approaches, communication styles, or problem-solving abilities that contributed to their own success. Alternatively, facial similarity could capture deeper compatibility between investors and founders that

facilitates productive working relationships. Under either interpretation, facial similarity would serve as a useful heuristic for identifying promising investments.

Conversely, if facial similarity reflects a bias that diverts investors from optimal decisions, we would expect negative performance effects. In this case, the tendency to fund facially similar founders would represent a costly deviation from merit-based evaluation.

Table V presents results from estimating equation 5, which tests these competing hypotheses. The dependent variable is successful exit via IPO or acquisition where the acquisition price exceeds twice the cumulative funding raised (Yimfor and Garfinkel, 2023). Column 1 shows that a one standard deviation increase in facial similarity is associated with a 0.212 percentage point decrease in exit probability. This negative effect persists across all specifications as we add controls.

[INSERT TABLE V ABOUT HERE.]

The coefficient remains stable as we progressively add controls for similar age (Column 2), shared network connections (Column 3), and the log of cumulative funding raised (Column 4). In the full specification, facial similarity continues to predict 0.200 percentage points lower exit probability. Given a baseline exit rate of 2.75% in our sample, this represents approximately a 7% decrease in the likelihood of successful exit.

The control variables reveal other patterns in venture outcomes. Same-gender pairs show higher exit rates (1.181 percentage points in Column 4), while shared network connections surprisingly predict lower exit rates (-0.800 percentage points). As expected, cumulative funding raised strongly predicts exits.

These results contradict the hypothesis that facial similarity captures valuable unobserved founder characteristics. Rather than serving as a proxy for cognitive ability, shared work styles, or other success-predicting traits, facial similarity appears to lead investors toward worse investment decisions. This finding suggests that appearance-based matching in venture capital represents a costly heuristic that reduces allocative efficiency.

5.7. Y Combinator Setting

To understand where in the investment process facial similarity operates, we exploit a unique feature of the Y Combinator ecosystem. Unlike traditional venture investing where investors both source and select deals, Y Combinator's accelerator

program separates these functions. Their competitive application process, which accepts less than 2% of applicants, provides external validation and standardized information about each startup. This setting allows us to test whether facial similarity influences final investment decisions when the sourcing and screening functions are performed by an intermediary.

Table VI presents results from the Y Combinator analysis. Panel A first confirms that demographic factors predict facial similarity in the YC sample similarly to the main sample. Same-race pairs show 0.173 standard deviations higher facial similarity, while same-gender pairs show 0.242 standard deviations higher similarity.

[INSERT TABLE VI ABOUT HERE.]

Panel B presents the key test from equation 6. Strikingly, facial similarity does not predict investment matching in the Y Combinator context. The coefficient on facial similarity is -0.146 and statistically insignificant (Column 5), in sharp contrast to the positive and significant effect in the general VC sample. This null result holds across all specifications.

Importantly, other matching patterns persist in the YC setting. Same-race pairs are 1.053 percentage points more likely to match, and shared network connections increase match probability by 3.397 percentage points. The persistence of these effects while facial similarity disappears suggests different mechanisms at work.

This pattern indicates that facial similarity functions primarily as a screening heuristic during initial deal sourcing rather than influencing final selection decisions. When Y Combinator's rigorous vetting process provides strong quality signals, investors no longer rely on facial appearance as a proxy for founder potential. The certification effect of YC acceptance appears to substitute for whatever informational value investors might otherwise extract from facial similarity. In contrast, demographic and network effects persist because they likely reflect deeper preferences or relationship dynamics that influence investment decisions regardless of external validation.

6. Conclusion

This paper documents a novel dimension of homophily in entrepreneurial finance: facial similarity between venture capital investors and startup founders. Using machine learning-based facial embeddings from 86,409 potential investor-founder pairs, we show that facial resemblance significantly predicts investment matching, even after controlling for demographic similarities and network connections. Three additional findings sharpen our understanding of this phenomenon.

First, the effect is concentrated among underrepresented investors—3.4 times larger for non-white investors and 2.3 times larger for female investors—suggesting that facial similarity may serve as a heuristic when traditional information channels are less accessible. Second, facial similarity predicts worse investment outcomes, with matched startups being 7% less likely to achieve successful exits. This indicates that appearance-based matching represents a costly bias rather than a valuable signal of founder quality. Third, facial similarity effects disappear entirely in the Y Combinator setting, where startups undergo external screening. This null result reveals that facial similarity operates primarily during initial deal sourcing rather than final selection decisions.

Together, these findings suggest that facial similarity functions as a screening heuristic that investors use to narrow their consideration sets, particularly when facing information constraints. While this heuristic may help investors manage deal flow, it appears to direct capital away from higher-quality investments. The concentration of these effects among underrepresented investors raises important questions about how information frictions in venture capital networks may perpetuate suboptimal matching patterns. Although we emphasize caution in causal interpretation, understanding these subtle influences on capital allocation is crucial for improving the efficiency and equity of entrepreneurial finance.

References

Bengtsson, O., Hsu, D.H., 2015. Ethnic matching in the u.s. venture capital market. Journal of Business Venturing 30, 338–354.

Bowles, D.C., McKone, E., Dawel, A., Duchaine, B., 2022. The relationship between cognitive style and facial processing: Evidence from a face matching task. Journal of Experimental Psychology: Human Perception and Performance 48, 27–41.

Brooks, A.W., Huang, L., Kearney, S.W., Murray, F.E., 2014. Investors prefer entrepreneurial ventures pitched by attractive men. Proceedings of the National Academy of Sciences 111, 4427–4431.

Cook, L.D., Marx, M., Yimfor, E., 2024. Funding Black High-Growth Startups. Working Paper 30682. Journal of Finance, Forthcoming.

Duarte, J., Siegel, S., Young, L., 2012. Trust and credit: The role of appearance in P2P lending. Review of Financial Studies 25, 2455–2484. doi:10.1093/rfs/hhs071.

Garfinkel, J.A., Mayer, E.J., Strebulaev, I.A., Yimfor, E., 2021. Alumni networks in venture capital financing. SMU Cox School of Business Research Paper .

Gigerenzer, G., Gaissmaier, W., 2011. Heuristic decision making. Annual review of psychology 62, 451–482.

Gompers, P., Huang, K., Wang, S., 2017. Homophily in Entrepreneurial Team Formation. Technical Report. Harvard Business School Working Paper.

Gompers, P.A., Gornall, W., Kaplan, S.N., Strebulaev, I.A., 2020a. How do venture capitalists make decisions? Journal of Financial Economics 135, 169–190.

Gompers, P.A., Gornall, W., Kaplan, S.N., Strebulaev, I.A., 2020b. How do venture capitalists make decisions? Journal of Financial Economics 135, 169–190.

Gompers, P.A., Mukharlyamov, V., Xuan, Y., 2016. The cost of friendship. Journal of Financial Economics 119, 626–644.

Guenzel, M., Kogan, S., Niessner, M., Shue, K., 2025. Ai personality extraction from faces: Labor market implications .

Huang, X., Ivković, Z., Jiang, J., Wang, I., 2023. Angel investment and first impressions. Journal of Financial Economics .

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47, 263–291.

Kosinski, M., 2021. Facial recognition technology can expose political orientation from naturalistic facial images. Scientific Reports 11, 1–7.

Todorov, A., Porter, J.M., 2015. Personality impressions from facial appearance. Oxford Handbook of Face Perception , 631–651.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. science 185, 1124–1131.

Yimfor, E., Garfinkel, J.A., 2023. Predicting success in entrepreneurial finance research. Journal of Corporate Finance , 102359.



Figure 1: FaceNet similarity among the Olsen sisters.

This figure demonstrates the effectiveness of FaceNet embeddings in capturing meaningful facial similarity within gender groups. The diagonal shows the standardized face images that FaceNet processes before generating embeddings. The similarity scores between pairs (displayed in off-diagonal positions) are calculated using cosine similarity between the 128-dimensional FaceNet embeddings, scaled to a 0-1 range. Higher scores indicate greater facial similarity as determined by the neural network. The gradient shows that FaceNet encodes detailed geometry rather than coarse gender or familial cues.

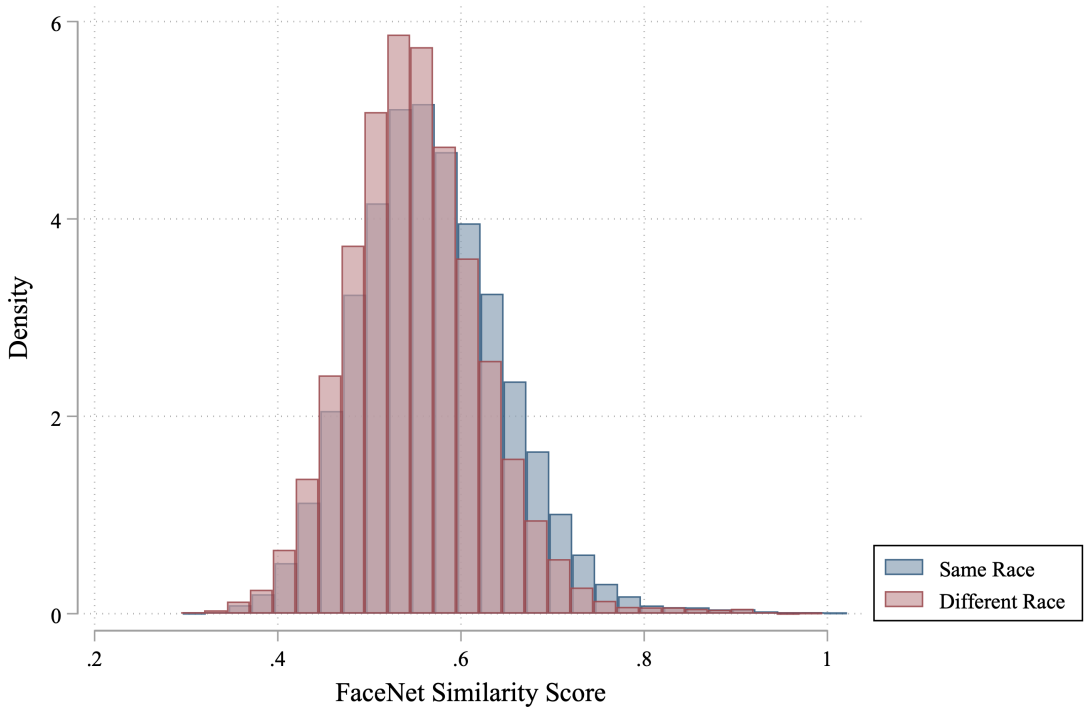
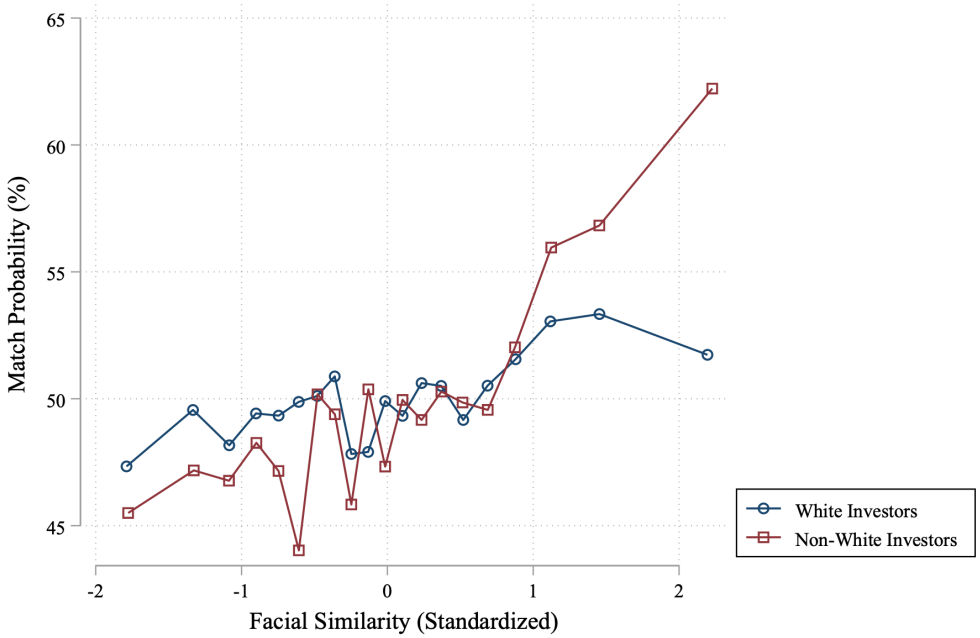


Figure 2: Distribution of Facial Similarity Scores Within and Across Racial Groups

This figure displays the distribution of FaceNet facial similarity scores between founder-investor pairs, separated by whether the individuals are of the same race (navy) or different races (maroon). The overlap between distributions demonstrates significant within-race variation in facial similarity, confirming that FaceNet captures meaningful facial feature differences beyond racial characteristics. Higher similarity scores indicate greater facial resemblance between pairs.

Panel A: Effect of Facial Similarity by Investor Race



Panel B: Effect of Facial Similarity by Investor Gender

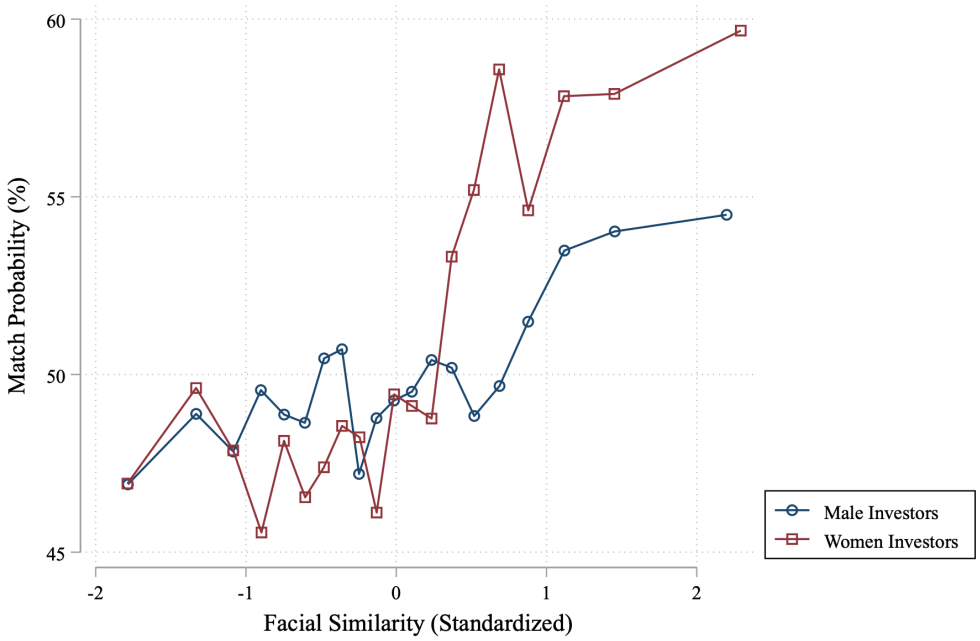


Figure 3: Facial Similarity and Investment Match Probability

This figure examines the relationship between facial similarity and venture capital investment decisions across investor demographics. Panel A presents a binned scatter plot showing how standardized facial similarity (FaceNet) between founders and investors affects the probability of an investment match, separately for white and non-white investors. Deals are grouped into bins along the horizontal axis based on facial similarity scores, with the vertical axis showing the percentage likelihood of a match. Panel B presents a similar analysis by investor gender, contrasting the effects for male versus female investors. Each point represents the average match probability for deals within each facial similarity bin.

Table I: Summary Statistics for Startups by Founder Race (2000-2020)

This table presents summary statistics comparing actual investment matches with counterfactual matches that could have happened but did not. Each observation represents a lead partner-founder pair, where the counterfactual pairs are constructed by matching on deal characteristics. Specifically, we match deals based on deal type, industry, founding year, and state, requiring deals to occur within a 180-day window, and selecting the closest match by deal size. This matching procedure creates a controlled comparison where the only difference between actual and counterfactual deals is the partner-founder pairing, allowing us to isolate homophily effects. The key variables are defined as follows: *Facial Similarity (Standardized)* is the standardized cosine similarity between partner and founder FaceNet embeddings (higher values indicate greater facial resemblance); *Same Race*, *Same Gender*, and *Similar Age* are indicators equal to one if the partner and founder share these demographic characteristics; *Shared Network Connection* is an indicator for whether the partner and founder attended the same university of worked at the same company before the deal year. *Non-White Investor* and *Female Investor* are indicators for partner demographics; *Deal Size* is the total capital raised in the financing round; *Angel Investment*, *Seed Round*, and *Early Stage VC* are indicators for deal type; industry indicators identify the startup's primary sector. The t-statistics test differences in means between actual and counterfactual deals. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. All continuous variables are winsorized at the 1% and 99% levels to minimize the influence of outliers.

	Actual Deal		**Counterfactual Deal**		
	N = 43,204		**N = 43,204**		
	Mean	Std. Dev.	Mean	Std. Dev.	t-stat
Facial Similarity (Standardized)	0.03	0.97	−0.04	0.95	10.77***
Same Race	0.65	0.48	0.60	0.49	15.73***
Same Gender	0.84	0.37	0.82	0.39	8.28***
Similar Age	0.34	0.47	0.31	0.46	8.74***
Shared Network Connection	0.07	0.26	0.03	0.17	28.59***
Non-White Investor	0.25	0.43	0.25	0.43	−0.01
Female Investor	0.09	0.29	0.09	0.29	0.00
Deal Size ($ Millions)	9.19	27.40	9.21	28.20	0.00
Deal Year	0.01	0.11	0.01	0.11	0.00
Angel Investment	0.46	0.50	0.46	0.50	0.00
Seed Round	0.53	0.50	0.53	0.50	0.00
Early Stage VC	2016.15	3.10	2016.15	3.10	0.00
Information Technology	0.13	0.33	0.13	0.33	0.00
Healthcare	0.19	0.39	0.19	0.39	0.00
Financial Services	0.00	0.06	0.00	0.06	0.00
Consumer Products & Services	0.02	0.14	0.02	0.14	0.00
Business Products & Services	0.09	0.29	0.09	0.29	0.00
Energy	0.56	0.50	0.56	0.50	0.00
Materials & Resources	0.00	0.04	0.00	0.04	0.00

Table II: Determinants of Facial Similarity Between Founders and Investors

This table presents results from our analysis of the demographic factors associated with facial similarity between founders and investors. We report coefficients from OLS regressions with standardized facial similarity (FaceNet) scores as the dependent variable. Standard errors are clustered at the partner level and shown in parentheses. Each column examines a different demographic factor that might relate to facial similarity: same race (Column 1), same gender (Column 2), similar age (Column 3), and all factors combined (Column 4). Year fixed effects are included in all specifications. Facial similarity is measured using FaceNet embeddings, which capture fine-grained facial features beyond demographic characteristics. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level.

	Dependent Variable: Facial Similarity (Standardized)			
	(1)	(2)	(3)	(4)
Same Race	0.228***			0.217***
	(0.009)			(0.009)
Same Gender		0.385***		0.371***
		(0.011)		(0.011)
Similar Age			0.138***	0.136***
			(0.009)	(0.009)
Observations	86,409	86,409	86,409	86,409
Adjusted-R^2	0.013	0.023	0.005	0.039
Year Fixed Effects	Yes	Yes	Yes	Yes

Table III: Effect of Facial Similarity on Investment Match Probability

This table examines how facial similarity between founders and investors affects the probability of investment matching. We report coefficients from OLS regressions where the dependent variable is an indicator for whether an investment deal occurred (multiplied by 100). Standard errors are clustered at the partner level and shown in parentheses. Column (1) shows the baseline effect of facial similarity, while subsequent columns add controls for demographic similarities: same race (Column 2), same gender (Column 3), similar age (Column 4), and shared network connections (Column 5). All specifications include year and partner fixed effects to account for temporal trends and time-invariant partner characteristics. Facial similarity is measured using standardized FaceNet embeddings, which capture fine-grained facial features that may influence investment decisions through homophily or other mechanisms. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level.

	Dependent Variable: Match Probability (%)				
	(1)	(2)	(3)	(4)	(5)
Facial Similarity (Standardized)	2.443***	1.967***	1.732***	1.639***	1.576***
	(0.232)	(0.230)	(0.232)	(0.231)	(0.229)
Same Race		7.611***	7.546***	7.528***	7.414***
		(0.551)	(0.552)	(0.551)	(0.547)
Same Gender			5.075***	5.060***	5.079***
			(0.754)	(0.754)	(0.748)
Similar Age				4.105***	3.826***
				(0.579)	(0.579)
Same Network					26.997***
					(0.998)
Observations	86,409	86,409	86,409	86,409	86,409
Adjusted-R^2	-0.164	-0.159	-0.158	-0.157	-0.144
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Partner Fixed Effects	Yes	Yes	Yes	Yes	Yes

Table IV: Heterogeneous Effects of Facial Similarity Across Investor Demographics

This table examines how the effect of facial similarity on investment matching varies across investor demographics and network connections. We report coefficients from OLS regressions with standard errors clustered at the partner level shown in parentheses. Column (1) shows the interaction between facial similarity and investor race, testing whether non-white investors respond differently to facial similarity than white investors. Column (2) shows the interaction with investor gender, testing for differential effects between male and female investors. Column (3) examines how shared network connections moderate the effect of facial similarity. Column (4) includes all interaction terms simultaneously to test their joint significance. All specifications include year and partner fixed effects. The dependent variable in all columns is Match Probability (percent), which indicates whether a real investment match occurred. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level.

	Dependent Variable: Match Probability (%)			
	(1)	(2)	(3)	(4)
Interaction:	Race	Gender	Network	All
Facial Similarity (Standardized)	0.844***	1.434***	1.469***	0.628**
	(0.264)	(0.236)	(0.235)	(0.271)
Facial Similarity × Non-White Investor	2.825***			2.766***
	(0.530)			(0.529)
Facial Similarity × Female Investor		1.935**		1.788*
		(0.932)		(0.929)
Facial Similarity × Shared Network			1.950**	1.820**
			(0.871)	(0.874)
Same Race	7.274***	7.419***	7.407***	7.276***
	(0.547)	(0.547)	(0.547)	(0.547)
Same Gender	5.145***	4.990***	5.078***	5.060***
	(0.748)	(0.744)	(0.748)	(0.745)
Similar Age	3.834***	3.829***	3.829***	3.838***
	(0.579)	(0.579)	(0.579)	(0.579)
Shared Network	26.979***	26.974***	26.904***	26.872***
	(0.995)	(0.998)	(1.001)	(0.998)
Observations	86,409	86,409	86,409	86,409
Adjusted-R^2	-0.14	-0.14	-0.14	-0.14
Year Fixed Effects	Yes	Yes	Yes	Yes
Partner Fixed Effects	Yes	Yes	Yes	Yes

Table V: Facial Similarity and Investment Outcomes

This table examines whether facial similarity between founders and investors predicts investment outcomes. We report coefficients from OLS regressions where the outcome measures is a successful exit (via an IPO or acquisition), where the acquisition deal price is at least twice cumulative amount of funding the startup raised. Specifications include controls for demographic similarities (same race, gender, age), network connections, and the cumulative amount of funding raised. Standard errors are clustered at the partner level and shown in parentheses. All specifications include year and partner fixed effects. The sample is restricted to deals from 2010-2018 to allow sufficient time for exits and follow-on financings to occur. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level.

	Dependent Variable: I(IPO/M&A) ×100			
	(1)	(2)	(3)	(4)
Facial Similarity (Standardized)	-0.212**	-0.206**	-0.205**	-0.200**
	(0.084)	(0.084)	(0.084)	(0.084)
Same Race	0.295*	0.294*	0.295*	0.292*
	(0.173)	(0.173)	(0.173)	(0.171)
Same Gender	1.346***	1.348***	1.350***	1.181***
	(0.160)	(0.160)	(0.160)	(0.157)
Similar Age		-0.217	-0.210	-0.216
		(0.180)	(0.180)	(0.180)
Shared Network			-0.709**	-0.800**
			(0.354)	(0.352)
Ln(Cumulative funding raised)				1.548***
				(0.087)
Observations	61,974	61,974	61,974	61,916
Adjusted-R^2	0.095	0.095	0.095	0.104
Year Fixed Effects	Yes	Yes	Yes	Yes
Partner Fixed Effects	Yes	Yes	Yes	Yes

Table VI: Facial Similarity in the Y Combinator Setting

This table examines facial similarity between Y Combinator founders and investors. Panel A presents results from OLS regressions with standardized facial similarity scores as the dependent variable, examining demographic factors that relate to facial similarity. Panel B examines how facial similarity affects the probability of investment matching following demo days, where the dependent variable is an indicator for whether an investment deal occurred (multiplied by 100) within 120 days after a company's Y Combinator demo day. Standard errors are clustered at the partner level and shown in parentheses. All specifications include year fixed effects, and Panel B additionally includes partner fixed effects. The analysis spans Y Combinator batches from 2010 to 2023. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level.

Panel A: Determinants of Facial Similarity

	Dependent Variable: Facial Similarity (Standardized)			
	(1)	(2)	(3)	(4)
Same Race	0.182***			0.173***
	(0.040)			(0.040)
Same Gender		0.256***		0.242***
		(0.055)		(0.058)
Similar Age			0.053	0.057
			(0.055)	(0.057)
Observations	15,128	15,128	15,128	15,128
Adjusted-R^2	0.010	0.011	0.003	0.019
Year Fixed Effects	Yes	Yes	Yes	Yes

Panel B: Effect on Investment Match Probability

	Dependent Variable: Match Probability (%)				
	(1)	(2)	(3)	(4)	(5)
Facial Similarity (Standardized)	-0.057	-0.115	-0.138	-0.140	-0.146
	(0.217)	(0.212)	(0.208)	(0.209)	(0.209)
Same Race		1.043**	1.018**	1.020**	1.053**
		(0.419)	(0.423)	(0.422)	(0.423)
Same Gender			0.836	0.843	0.902
			(0.663)	(0.662)	(0.664)
Similar Age				0.809	0.833
				(0.603)	(0.599)
Shared Network					3.397***
					(1.271)
Observations	15,127	15,127	15,127	15,127	15,127
Adjusted-R^2	0.025	0.025	0.025	0.025	0.026
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Partner Fixed Effects	Yes	Yes	Yes	Yes	Yes