Preferences in Social Network Formation*

Celina Högn†, Markus Nagler‡ & Johannes Rincke**

February 14, 2025

Preliminary draft. Please do not cite or circulate without permission.

Abstract

Social networks are typically characterized by strong segmentation regarding individual characteristics. We study to what extent homophily in social network formation is driven by preferences for peers carrying similar characteristics, as opposed to common preferences for specific characteristics or common environments. Combining a stated-choice experiment with administrative and survey data, we describe homophily in an emerging network of university students and separately uncover students' underlying preferences. We find that the strong homophily observed in our data cannot be explained by homophilous preferences. Rather, students have common valuations, making some students generally more sought-after network partners.

*We thank Martin Abraham, Katrin Auspurg, Thomas Buser, Christina Felfe, Ulrich Glogowsky, Fabian Kosse, Matthew Lindquist, Malte Reichelt, Ulf Zölitz, and seminar participants at SOFI Stockholm as well as conference participants at various conferences for helpful comments and suggestions. We are grateful to the German Science Foundation for funding under the grants NA 1722/1-1 and RI 1959/4-1. Nagler thanks the Universitätsbund Erlangen-Nuremberg for financial support. We pre-registered the experimental design at the AEA RCT Registry (https://doi.org/10.1257/rct.12348-1.0).

†University of Erlangen-Nuremberg; celina.hoegn@fau.de

‡University of Erlangen-Nuremberg, CESifo, IZA, and LASER; markus.nagler@fau.de

**University of Erlangen-Nuremberg and CESifo; johannes.rincke@fau.de

1 Introduction

Networks play a key role in shaping individual outcomes. For example, recent research shows that personal networks influence educational achievements, job aspirations, labor market referrals, beliefs and personality (e.g., Zeltzer, 2020; Beaman et al., 2018; Norris, 2020; Golsteyn et al., 2021; Shan and Zölitz, 2022; Bailey et al., 2024). A common feature of these networks is homophily, i.e., the tendency for individuals with similar characteristics to connect more frequently (e.g., McPherson et al., 2001; Jackson, 2010).

While key features and consequences of social networks have been documented, the forces underlying their structure are less understood. This is particularly true for the role of individual preferences for network partners. Homophily could arise through preferences or opportunity (Feld and Grofman, 2011).[1] Specifically, it could arise through homophilous preferences for like-minded peers (i.e., preferences to interact with somebody of one's own type), through common preferences for specific characteristics or traits (i.e., preferences to interact with somebody of a specific type irrespective of one's own type), or through proximity (i.e., due to self-selection on common factors into environments), without these explanations being mutually exclusive (Jackson, 2014).[2] Theoretical models of network formation often assume homophilous preferences, but there is no direct evidence on this (e.g., Zuckerman, 2024; Currarini et al., 2009). Understanding the sources of homophily in networks is also relevant since some offer scope for interventions that could help, for example, to reduce inequalities or discrimination (e.g., Jackson, 2014; Small and Pager, 2020).

This paper provides direct evidence on the role of preferences for homophily in social network formation. The emerging social network we study is one among undergraduate students at a large public university in Germany. Our setting allows to explore homophily in networks in terms of students' characteristics and personality. We leverage a combination of administrative data, survey data, and a stated-choice experiment among one cohort of undergraduate students to shed new light on the sources of homophily in personal networks.

Our empirical analysis proceeds in two steps. We first describe and characterize the existing social network among this cohort after one year at the university. To this end,

[1] In principle, homophily in malleable characteristics could also arise due to reverse causality, say, if individuals who interact become more similar over time (e.g., Feld and Grofman, 2011). This is impossible for dimensions that are not malleable and is not an issue in our context because we analyze the formation of a social network.

[2] For example, consider homophily in a trait such as altruism. One could observe such homophily because similarly altruistic individuals self-select into similar environments and therefore connect at higher rates, because less altruistic individuals like to connect with less altruistic individuals while more altruistic individuals like to connect to their like, or because *everybody* wants to connect with more altruistic individuals (but more altruistic individuals do not want to connect to less altruistic individuals).

we link rich administrative data on the students' characteristics (like gender, secondary GPA, field of study, etc.) with two waves of survey data. The first wave was conducted around one week after students started studying and covers, e.g., a wide range of personality traits. The second wave was conducted after one year at the university. In this survey wave, we asked the students about their closest contacts among their fellow students. From this data, we identify the networks of 416 individuals, resulting in a dataset covering around 1,300 social ties.

Using this data, we show that students in our setting exhibit strong homophily on observable characteristics and behavioral traits. For most of our measures, we find that common characteristics strongly predict network connections. The largest effect is a shared field of study, leading to almost a fivefold increase in the likelihood to connect. We also find homophily in gender and in speaking the same second mother tongue besides German (both more or less doubling the likelihood), in having a similar secondary GPA ($> 50\%$ increase in likelihood), in a shared primary study motive, and in having a similar personality measured by conscientiousness, altruism, competitiveness, and self-efficacy. These results extend and are in line with the existing literature on networks which finds homophily on observable characteristics and on some measures of personality (Jackson et al., 2017; Bhargava et al., 2022).

The second step of our analysis directly elicits students' preferences for network partners. To this end, we ran an experiment among our cohort in the first week of their studies, directly after their participation in the first survey wave. We administered an incentivized stated-choice experiment about network partners. In the experiment, the 1,494 participants evaluated a set of fictitious profiles of other students. For each profile, we elicited whether the respondent would be interested to connect to a student with a respective profile. Each profile described a fictitious student in terms of study field, gender, second mother tongue, primary study motivation, secondary GPA, and personality traits comprising conscientiousness, altruism, competitiveness, and self-efficacy. The attribute values were randomly drawn.[3] To encourage truthful responses, participants could (voluntarily) participate in an online networking event to meet other students based on their stated preferences for network partners.

Leveraging the data from the stated-choice experiment, we find that homophilous preferences are the exception rather than the norm and cannot explain the strong homophily in the existing network found in the first step of our analysis. Across most characteristics and traits, the experimental data reveal common preferences.

[3]We ensured that the profiles would be realistic and the distribution of characteristics in line with actual distributions. For instance, to account for the fact that most of the actual social ties form among students of the same field of study, we presented each participant with profiles where the likelihood of a shared field of study between participant and profile was 60%, and all remaining fields of study together appeared with a likelihood of 40%.

That is, irrespective of their own type in the respective dimension, individuals on average prefer a specific set of characteristics. This is true in particular for altruism and conscientiousness: Highly altruistic or highly conscientious individuals are more popular potential network partners irrespective of the respondent's characteristics. Similarly, we find common preferences for female contacts, contacts with high self-efficacy, and contacts whose main study motive is to have an impact on society. There is not a single dimension in which students show clear homophilous preferences in the sense that each type of student significantly prefers their own type relative to the opposite type. Students also show no "sorting" into network connections in the sense that preferences for specific characteristics is significantly higher among those students that have friends with such characteristics. We validate these preference estimates by leveraging the actual characteristics of students who participate in the follow-up survey after one year. We show that students with high estimated likelihood of being accepted as network partners have significantly larger networks at university.[4]

Our paper primarily contributes to the literature on network formation. There, homophily has been identified as a key feature of networks.[5] Homophily has for example been identified regarding race, gender, income, and religion (e.g., Bailey et al., 2018; Chetty et al., 2022).[6] Only few papers consider homophily in behavioral traits. Closest to our study is Bhargava et al. (2022), who study French high school students and find evidence for homophily based on demographics as well as on all elicited behavioral traits. In their data, homophily is especially pronounced in morality, depth of reasoning, cooperation, and generosity. Jackson et al. (2023) observe student networks at Caltech over three years and find evidence for homophily in gender and ethnicity across all layers, as well as some homophily on malleable characteristics.[7]

Relative to the established literature identifying homophily in social networks, there is less work on the sources of homophily. One important driver of homophily is exposure, due to contact in shared spaces of people with similar characteristics and interests (e.g., Jackson, 2021; Jackson et al., 2017; Kossinets and Watts, 2009; Feld and Grofman, 2011; Kruse et al., 2016; Tekles et al., 2022). Homophily can, however, also be the result of active choices about whom to interact with (Jackson, 2021). Theoretical

[4]To illustrate the relevance this outcome, in our data larger networks at university predict higher life satisfaction, lower depression and loneliness scores, and higher satisfaction with current networks and network sizes.

[5]Homophily has also been found to influence economic outcomes like information diffusion, job opportunities, mobility, marriage markets, health behavior, and education (for a review, see Jackson et al., 2017).

[6]See also Girard et al. (2015) for evidence on homophily in social networks among German undergraduates.

[7]The literature has also studied mating behavior. Multiple studies show that females have a preference for highly educated males, while men prefer physically attractive females. Some studies also find that men do not value women's intelligence or ambition when it exceeds their own (e.g., Fisman et al., 2006; Hitsch et al., 2010; Skopek et al., 2011; Ong, 2016; Neyt et al., 2019; Shan and Zölitz, 2022).

models of network formation in economics often assume homophilous preferences (e.g., Zuckerman, 2024; Currarini et al., 2009). In sociology, researchers have used panel data to identify the role of what is there referred to as "choice-based homophily" (see, e.g., Smith et al., 2014a; Melamed et al., 2020; Weber et al., 2020).[8] Homophilous preferences may play a role here, but common preferences for specific characteristics are prevalent (e.g., Abele and Brack, 2013) and may also result in homophily since, in economics language, there is strong demand for friendship with specific potential network partners, but little supply (e.g., Wimmer and Lewis, 2010; Schaefer et al., 2011). A rare exception is Snijders and Lomi (2019), who use longitudinal data to estimate patterns consistent with choice-based homophily beyond homophilous preferences.[9] We are not aware of any paper directly and causally estimating these preferences. Our main contribution to this literature therefore is a clean, direct, and causal experimental approach capable of uncovering to what extent the observed homophily in an emerging social network can be due to homophilous preferences. We expand the literature by showing common rather than homophilous preferences for most characteristics and personality traits, including dimensions that have not been studied before, like conscientiousness, self-efficacy, and measures of study motivation.

Our paper also contributes to the literature on peer effects in education. This literature has provided some evidence for an important role of peers. In a vast literature on peer effects in education, for example, peers seem to affect educational attainment (Sacerdote, 2014) and prosocial behavior (Alan et al., 2021; Rao, 2019). Bhargava et al. (2022) provide a comprehensive literature review.[10] We add to this literature by providing evidence on the fundamental sources of peer group formation, in particular on the role of (homophilous) preferences.

Finally, a literature in sociology and psychology investigates how subjects value other subjects' traits, often in hypothetical scenarios (e.g., Abele and Brack, 2013; Porter and Rigby, 2019; Abele et al., 2020; Ollroge and Sawert, 2022). To our knowledge, there is no paper, however, that disentangles these preferences. We thus add to this research by explicitly testing for homophilous preferences. We also add by incentivizing our stated choice experiment, providing students with actual network partners based on their choices. We finally add by combining our experiment with administrative and survey data which allows us to show actual homophily in student networks. This unique data link also allows us to investigate sorting into our experiment.

[8] In the sociological network dynamics literature, researchers have primarily focused on friendship formation in (early) adolescence and on longitudinal data to disentangle choice-based and environment-based homophily (see, e.g., Knecht et al., 2010; Flashman, 2011; Lomi et al., 2011; Sawka et al., 2013).

[9] More recently, see, e.g., Schaefer and Kreager (2020) for related work.

[10] For peer effects in networks, see also Shkoza et al. (2020) and Bramoullé et al. (2020).

2 Data Sets

The central element in our analysis aims at identifying preferences for network partners. To this end, we use data on students belonging to the cohort starting their studies at one large public university in Germany in fall 2023 ($N = 3{,}572$). Our data combines three different sources. First, for students' background characteristics we rely on administrative data from the university. Second, all students belonging to the cohort were invited to a survey just after commencing their studies in fall 2023. In this survey, we elicited a comprehensive set of measures of personality traits for a subset of 1,626 students.[11] Third, all students who had completed the survey were invited to participate in a stated-choice experiment aimed at eliciting preferences for network partners right after the completion of the survey on traits. Of all students in the cohort, 1,494 completed both survey modules, providing us with complete information on the respective respondent's own personality traits and choice data regarding potential network partners.[12] In summary, the data we use to identify preferences for network partners comprises 1,494 individuals from the cohort 2023. This data covers individual background characteristics, personality traits, and stated-choice data regarding potential network partners.

Table 1 shows descriptive statistics for the sample of students who participated in the stated-choice experiment.[13] Table A1 in the Appendix analyzes sorting into the experiment. Students participating in the experiment are very similar to the student population in terms of gender and age, but have slightly better high school GPAs. Our sample comprises students from all study fields, with the shares in our sample being close to overall shares for most fields.

In July 2024, we invited all students who had participated in the 2023 cohort survey for a follow-up survey. In this second survey wave, participants were asked to name their closest contacts at the university.[14] 570 individuals participated in the follow-up

[11]This initial cohort survey was part of a larger research agenda on students' traits, economic preferences, and study motives (Adler et al., 2024). The personality traits are based on well-established and validated survey items. Specifically, we elicit conscientiousness using a German version of the Big-5 survey kit (Gerlitz and Schupp, 2005). To measure self-efficacy (i.e., the expectation to succeed in certain situations or subjects based on one's own competencies and abilities), we rely on a similarly well-established scale (Beierlein et al., 2012). Our measure for altruism stems from the German version of the Preference Survey Module of the Global Preference Survey (Falk et al., 2018, 2023). Finally, we rely on a widely used method to elicit the willingness to compete (Niederle and Vesterlund, 2007; Buser et al., forthcoming).

[12]Attrition between survey modules was low: Of all students who took the survey on traits, only about 8% opted not to participate in the stated-choice experiment.

[13]For details regarding the experimental design and the choice data elicited during the experiment, please see Section 4.

[14]Specifically, we asked (translated from German): "We would like to know how you are connected at the university. Please use the following fields to list up to 5 of your closest contacts. Please provide the first and last names of your friends, as well as their field of study. If you want to list more friends,

Table 1: Descriptive Statistics of Experimental Sample

	Mean (1)	SD (2)	p10 (3)	p90 (4)
Female	0.57	0.50	0.0	1.0
Age	19.02	2.44	17.0	21.0
Secondary GPA	2.83	0.67	1.9	3.8
Has second mother tongue	0.29	0.45	0.0	1.0
Study motive: Interest in Subject	3.70	0.55	3.0	4.0
Study motive: Career	3.20	0.86	2.0	4.0
Study motive: Impact on Society	2.89	0.96	2.0	4.0

Note: This table provides descriptive statistics for the 1,494 students who participated in our experiment. The study motives are measured on a scale of 1 to 4, with 4 indicating the highest importance when selecting the field of study (selecting multiple motives was possible). See Section 3.1 for more details on our elicitation methods. In addition to the displayed variables, we use measures of personality traits. Since these variables are standardized, we do not display summary statistics for them.

survey. Table A2 shows that these individuals again on average have slightly better high school GPAs and there are again slight differences in the field of study. Out of the 570 students, 416 individuals provided information on at least one friend that we could identify in the administrative data. As seen in Table A2, apart from students without network data being slightly older than others, there are no significant differences between those who named a friend and those who did not. A total of 1,009 individuals from the 2023 cohort were named as a friend at least once, with some students being named more often. Overall, the elicitation of friends allows us to identify 1,317 binary friendship relations.

Information on characteristics and traits is complete for all of the 416 participants in the follow-up survey who provided information on their network. This is because we invited only students who had completed the initial survey in 2023 to that survey. Regarding data on those students who were named as friends in the follow-up survey, information on background characteristics (obtained from the administrative data) is also complete. However, information on personality traits is incomplete, since not all students named as friends participated in the initial cohort survey in 2023 when traits were elicited. Overall, our data contain personality traits for 472 out of the 1,009 students who were named as friends. As a result, we can completely describe (in terms of characteristics and traits) around 700 friendship connections.[15]

you can do so on the next page." We did not restrict the number of answers to avoid censoring problems (Griffith, 2022).

[15]In addition to this data, we also elicited student networks in the students cohorts that started their degrees in 2020 and 2021. We describe these cohorts and show homophily in all dimensions in Appendix A.3. Instead of eliciting networks one year after the beginning of the students' studies as in our main cohort, we elicited the networks after nearly three years of enrollment in university. All student networks show consistent patterns of homophily. The only noticeable difference is that homophily in characteristics such as gender and second mother tongue is somewhat less pronounced (but still very large) in the networks that have formed over a longer time period. This could be due to those characteristics being more easily observed initially, such that these characteristics play a larger

In the following, we use the data on friendship connections to describe homophily in the social networks of students belonging to the 2023 cohort. When analyzing homophily in characteristics, we use all data on friendship connections we have information on. When studying homophily in personality traits, we use the smaller sample of friendship connections between students we can completely describe in terms of characteristics and personality traits.

3 Homophily in Characteristics and Personality

3.1 Empirical Approach

We now present our approach to identify homophily in an existing network of students, using the data on the cohort who enrolled in 2020. To assess homophily, we follow Bhargava et al. (2022) and estimate, separately for each dimension (characteristic or trait) k, one of the following equations:

$$connect_{ij} = \beta + \gamma^k \mathbb{1}[x_j^k = x_i^k] + \xi_i + \psi_j + \epsilon_{ij} \tag{1}$$

$$connect_{ij} = \beta + \gamma^k (-|y_i^k - y_j^k|) + \xi_i + \psi_j + \epsilon_{ij} \tag{2}$$

Equation (1) refers to binary dimensions k (like gender). $connect_{ij}$ is an indicator for a directed friendship pair consisting of subjects i and j, x_i^k and x_j^k are the (binary) characteristics of i and j in dimension k, $\mathbb{1}[\cdot]$ is the indicator function, and ξ_i and ψ_j are sender and receiver fixed effects. If k is a non-binary dimensions (like a trait or GPA), we estimate Equation (2), where $-|y_i - y_j|$ measures the closeness in standardized traits of subjects i and j. The coefficient measuring homophily is γ. In (1), it gives the change in the likelihood of observing a connection between students i and j if they share a given characteristic, controlling for sender and receiver fixed effects. In (2), γ measures the change in the likelihood of observing a connection between students i and j in response to a reduction in both subjects' distance in terms of k that is equivalent to one standard-deviation of the underlying (standardized) traits. Note that we allow connections to be unilateral (i.e., $connect_{ij}$ takes value 1 if either i or j state to be connected to the other subject, or both). We allow for the clustering of standard errors at the subject level. We report coefficients relative to the baseline probability for two subjects in our data to be connected, which is 0.20% for the sample covering only demographic characteristics and 0.22% for the sample also including traits.

role for network formation. Speaking against such an interpretation, homophily in these dimensions is still very strong in networks measured after three years. In addition, it does not seem like homophily in personality traits is more pronounced in the networks measured after three years than in those measured after one year.

3.2 Homophily is Prevalent in Student Networks

Figure 1 reports estimates of γ, the homophily coefficient. It shows that there is strong homophily in observable characteristics and student personality in our sample. Starting with the upper part of the figure, having the same gender nearly doubles the likelihood of observing a connection relative to the baseline likelihood, conditional on the characteristics of the students. There is also strong homophily in speaking a second language other than German. The figure also reveals homophily in the main motive for choosing one's field of study (interest in the field, having a successful career, and having an impact on society). We are not aware of other papers showing homophily in students' study motives.

In the bottom part of the figure, we show that there is strong homophily regarding secondary GPA as well. Having a similar GPA increases the likelihood of being connected by 60%. There is also strong homophily in student personality, namely in altruism, conscientiousness, self-efficacy, and competitiveness.[16] The evidence from Figure 1 on traits is in line with and extends recent findings by Bhargava et al. (2022), who show homophily in some dimensions of student personality in French schools.[17] In unreported regressions, we find that homophily is even larger in the dimension of field of study. Relative to the baseline probability of being connected, sharing the same field of study nearly leads to a fivefold increase in the likelihood of a social tie. Interestingly, it seems like common easily observable characteristics of fellow students such as gender, speaking a second language or field of study increase the likelihood of being connected more than less easily observable characteristics such as study motives.

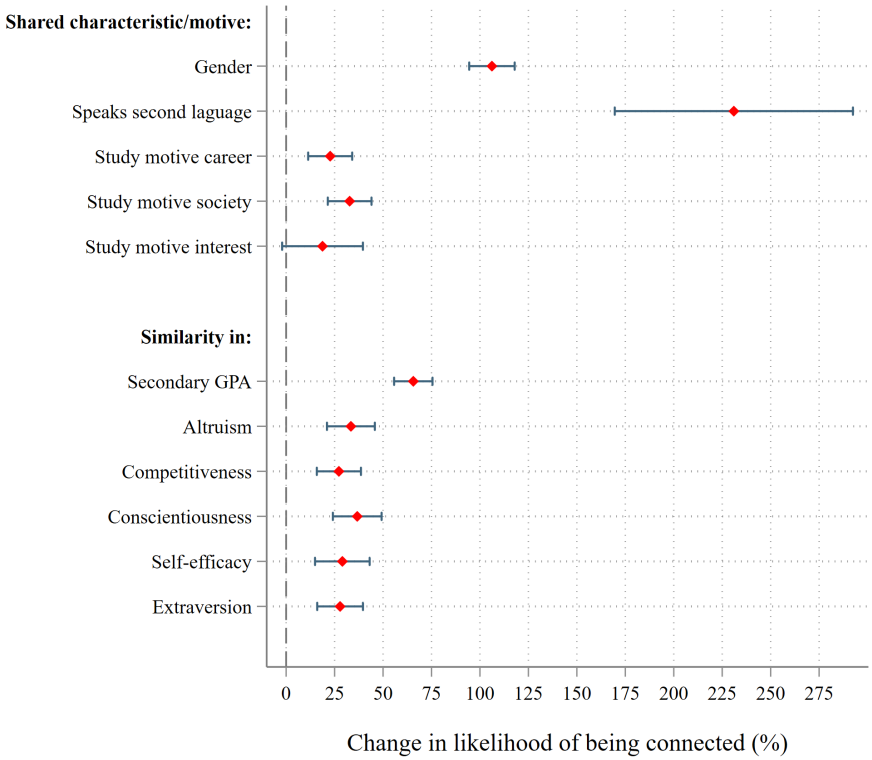
All in all, these findings are in line with and extend recent literature on homophily in emerging networks. Whereas patterns of homophily are common and have been shown to exist in various settings, the sources of homophily are not well understood (Feld and Grofman, 2011). For example, besides sorting into commong environments, the large homophily we observe in sharing a second language besides German could stem both from discrimination or from homophily (Leszczensky and Pink, 2019). Thus, disentangling the various possible explanations for why individuals tend to be socially connected to peers with similar characteristics is relevant: Depending on the source of homophily, attempts to alter the structure of social networks (e.g., in order to reduce inequality or mitigate discrimination) are more or less likely to succeed.[18] The natural

[16]Note, however, that we have fewer observations for these regressions since they rely on information only available for the subset of students who participated in the initial cohort survey in 2023. See Section 2 for details.

[17]See Appendix Figure A1 for estimates of homophily in additional dimensions of personality.

[18]For example, Carrell et al. (2013) find that exogenously manipulating peer groups in higher education did not improve outcomes because of endogenous peer group formation at the micro level. Recently, Elwert et al. (2023) found that more interethnic exposure does not affect discriminatory attitudes.

Figure 1: Observational Approach: Homophily in Student Networks



Note: This figure is based on observational data and shows homophily in demographic characteristics, study motives, and traits. Each coefficient measures homophily in the respective dimension and corresponds to a separate regression based on equation (1) or (2), depending on whether or not the respective dimension is binary. For gender and GPA, the sample consists of all (potential) connections between the 416 subjects naming at least one contact and the 1,009 subjects named at least once as friends. For all other dimensions, the sample consists of all (potential) connections between individuals and the 472 subjects named as friends who took part in the initial cohort survey in 2023. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

next step in our analysis is therefore to isolate the role of preferences as a possible source behind the observed patterns of homophily in our data.

4 Preferences for Potential Network Partners

4.1 Experimental Design and Empirical Approach

To estimate students' preferences over potential network partners, we exploit a data set that links administrative data on individual characteristics with survey data (including personality traits) and the data from a stated-choice experiment. Both the survey and the experimental data were elicited via the online survey administered to first-year students at our university in the fall of 2023 (see Section 2 for details). The experiment took place shortly after students entered university, implying that we study preferences in an emerging social network that are not influenced by students'

experiences or existing networks at university. In the following, we describe the experimental procedures.

The stated-choice experiment was administered after the elicitation of personality traits.[19] Importantly, when answering the survey questions on traits, participants did not know about the survey module containing the stated-choice experiment and the option to later participate in a networking event. This prevents the elicitation of personality traits to be distorted by considerations regarding the chances to be matched with other students for networking.

The instructions at the beginning of the experiment informed participants that we would show them fictitious profiles of other students and asked participants to state whether they would be interested in connecting with such students. The participants were informed that their stated preferences over potential network partners would be used to match them with other students during an online networking event. We made clear that participation in this event was voluntary.[20] As a result, decisions in the experiment had (at least potentially) real consequences in terms of networking options.[21] Apart from the networking incentive, participants earned a fixed reward of €4 for completing the stated-choice experiment. This payoff came on top of the fixed reward of €15 for completing the survey on personality traits and study motives. The instructions also asked participants to assume that the profiles were identical in dimensions not listed. We did so to insure that students would not be worried about correlated unobservables (Smith et al., 2014b; Gallen and Wasserman, forthcoming).[22] Overall, 1,494 students completed the experiment.

The experiment consisted of three screens each showing three different fictitious profiles of other students as potential network partners. The fictitious profiles described potential network partners in nine dimensions, including gender, second mother tongue other than German (none/Arabian/Spanish/Turkish), above-average GPA (yes/no), and study field (engineering, teaching, natural sciences, medicine, law, social sciences, arts and humanities). In addition, the profiles contained information on the main study motive (having a good career/interest in the subject/having an impact on society) and a set of indicators capturing personality traits, namely altruism, conscientiousness, competitiveness, and self-efficacy. These indicators described the

[19]We pre-registered the experiment in the AEA Registry as AEARCTR-0012348, see `https://doi.org/10.1257/rct.12348-1.0`.

[20]The networking event was held at the beginning of the participants' second term.

[21]Our results are qualitatively identical if we restrict the sample to participants who indicated that they would be interested in the online networking event (see Online Appendix Figure A11).

[22]Note that students who intend to participate in the networking event might still worry about correlated unobservables. For example, students from a specific field of study might differ in economic preferences and behavioral traits in dimensions other than the ones displayed (Adler et al., 2024). In subsamples that drop students who intend to participate in the networking event, we find that our results are qualitatively identical. We discuss this in Section 4.2.

profiles as "very altruistic" (yes/no), "very reliable and conscientious" (yes/no), "very competitive" (yes/no), and "trusts heavily in own abilities" (yes/no). For each profile, participants had to decide whether to accept or reject the potential network partner (yes/no/maybe).[23] Figure A6 in the Online Appendix shows a sample choice screen.

For each profile, all attribute values were randomly drawn. Between profiles, the draws were independent. Within each dimension, the attribute values were drawn with equal probabilities, with two exceptions: First, to account for the fact that most actual social ties form among students of the same field of study, we presented each participant with profiles where the likelihood of a shared field of study between participant and profile was 60%. The remaining 40% were equally split between all remaining fields of study. Second, to achieve a distribution of profiles regarding a second mother tongue other than German that would be similar to the actual distribution in our sample, we set the probability of a profile not having any second mother tongue to 80%. The remaining 20% were equally split between the three most frequent languages (as second mother tongue) in our sample (Arabian/Spanish/Turkish). Furthermore, to avoid outlier profiles, we ensured that for at least two personality traits, the attribute value "yes" was drawn in each profile. We also made sure that for any profile, the attribute value "yes" was never shown for all four personality traits.[24]

To address the concern that students may perceive the experiment as a dating opportunity (e.g., Fisman et al., 2006), we excluded gender from the profile attributes on the last choice screen (i.e., for the last three profiles). In all other dimensions, the attributes were the same as on the first two screens. Our main results are based on the subsample of evaluations of complete profiles (including gender), however.

The goal of the experiment is to allow for causal inference on the degree of homophily in preferences for network partners. Specifically, we ask to what extent the observed homophily in the actual network of students can stem from students' demand for network partners that have characteristics similar to their own. Alternatively, homophily in a social network could also be due to common preferences (i.e., a situation where individuals prefer network partners with a certain characteristic or trait, irrespective of whether or not they themselves carry this characteristic or trait). For example, it could be that all students in our sample prefer to connect to very altruistic peers. Very altruistic students, given their own preference and the demand from all others to form social ties with them, would then be able to connect to

[23]We added the *maybe* option to mitigate concerns that replies could be (partly) driven by social desirability. In our main analyses, we code *maybe*-replies as indicating rejections. Our results are robust to coding these replies as acceptances.

[24]We implemented this by re-drawing all personality traits in case of initial draws not fulfilling these constraints. Finally, to maximize power, we re-drew attribute values in cases where, for a given participant, two or more profiles would have been identical.

other altruistic students and reject social ties proposed by less altruistic peers. As a result, less altruistic students would have no choice but to connect among each other. Observationally, despite common preferences to connect with altruistic peers, this would result in a network structure characterized by homophily.

To determine the degree of homophily in preferences for network partners, we estimate

$$accept_{ij} = c + \sum_{k=1}^{K} \left[\delta^k (x_j^k = 1) + \eta^k (x_i^k = 1) + \theta^k (x_j^k = 1)(x_i^k = 1) \right] + \epsilon_{ij}, \qquad (3)$$

where $accept_{ij}$ indicates whether respondent i prefers to form a social tie with a peer characterized by profile j, x_i^k is the respondent's characteristic in dimension k, and x_j^k is the profile characteristic in dimension k. Thus, in our estimation, δ^k captures the preference for attribute k of profile j if $x_i^k = 0$ (i.e., the preference for a specific characteristic of a fictitious profile when the respondent does not share this characteristic). Similarly, $\delta^k + \theta^k$ captures the preference for attribute k of profile j if $x_i^k = 1$. For completeness, η^k measures how being a bearer of characteristic k affects the likelihood to prefer a connection if $x_j^k = 0$.

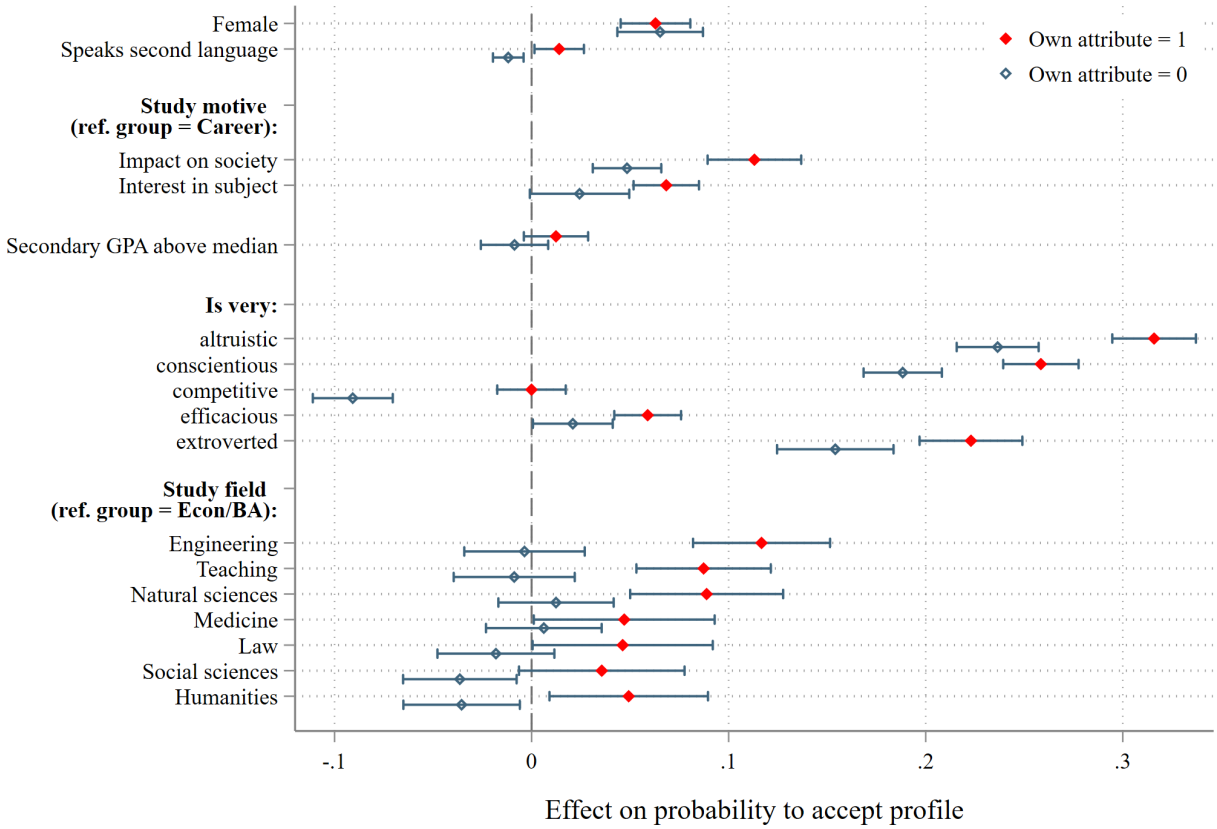
We characterize preferences as homophilous if bearers of a trait prefer other bearers of a trait, while non-bearers prefer non-bearers. In contrast, students have common preferences when bearers of a trait and non-bearers both prefer bearers or non-bearers of a trait. Finally, students have heterophilous preferences if bearers of a trait prefer non-bearers and vice-versa. In terms of interpretation, Figure A7 in the Online Appendix shows, separately for each type of preferences, a fictitious example of how parameter estimates indicating the respective preferences would look like.

4.2 Student Preferences for Network Partners

Figure 2 reports the results from estimating equation (3) using our experimental data. For each dimension, the blue diamond represents the estimate of δ^k (i.e., the preference for profile characteristic k among respondents who do not share this characteristic). The corresponding red diamond shows the estimate of $\delta^k + \theta^k$ (i.e., the preference for profile characteristic k among respondents who are bearers of the respective characteristic).[25] The figure also shows 95% confidence intervals based on standard errors that allow for clustering at the respondent level. The sample consists of all decisions made by the 1,494 participants when considering complete profiles (including gender). With each participant evaluating six of these profiles (two screens showing three profiles each), the sample comprises $1{,}494 \times 6 = 8{,}964$ observations. Across all choices, the

[25] In Appendix A.6, we show that alternative plausible approaches would not be able to identify homophilous preferences.

Figure 2: Choosing Network Partners: Common Rather Than Homophilous Preferences



Note: This figure shows preferences for network partners by the potential partner's demographic characteristics, study motives, personality traits and field of study. Based on equation (3), coefficients represent changes in the probability of a profile being selected as a potential network partner. The red estimates indicate preferences of respondents that hold the respective attribute. Blue diamonds denote preferences of respondents who do not hold the respective attribute. The dependent variable is an indicator equal to 1 if respondent i accepts a profile j as a potential partner and 0 otherwise. The sample consists of all decisions made by the 1,494 participants when considering profiles including gender, and thus $1494 \times 6 = 8964$ observations. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

probability for subjects accepting a profile as representing a potential network partner was 46.0%.

Figure 2 reveals that in our setting, if at all present, homophilous preferences are the exception rather than the norm. In several of the dimensions, students rather have common preferences for the respective characteristics. This holds true regarding the preference to connect with female students, with students with above-median altruism, conscientiousness, and self-efficacy, and with students whose main reason for choosing their field of study was to have an impact on society. For instance, among very altruistic subjects, profiles that show above-median altruism are 34 percentage points more likely to be selected as potential network partners. The point estimate for subjects with below-median altruism is a bit lower, but still positive and highly significant, indicating a similar preference for very altruistic network partners.

Leaving statistical significance aside, students who are above-median competitive (weakly) prefer to form social ties with other students who are particularly competitive (p-value $= 0.177$), while students who have below-median competitiveness clearly prefer network partners who are similar in that respect. We also see somewhat homophilous preferences for an above-average GPA in high school and for speaking a second language, but for some of the coefficient estimates, the confidence intervals again include zero.[26]

Figure 2 also shows that students have some preferences to connect to other students from their own field of study (relative to students in Business and Economics), but the estimates are relatively small and often insignificant. Students from some fields of study seem unpopular, with students outside the fields of Social Sciences and Humanities preferring not to interact with students from these fields, all else being equal. Our results suggest clear homophilous preferences only in the case of Humanities. We acknowledge, however, that one could interpret the preference of students to interact with their own field while not caring about other (business and economics) students as homophily. This is the case in Engineering, Teaching, and the Natural Sciences.

All in all, however, the preference estimates in all dimensions differ markedly from the strong homophily observed in Figure 1. We conclude that in most dimensions in our analysis, the homophily in the existing network of students cannot be explained by homophilous preferences.

Validation of preference estimates. To validate our preference estimates we assess whether the estimates have predictive value in actual student networks. To do so, we leverage our data on the personal characteristics of all respondents that provided information on their networks in the follow-up survey after one year at the university. From these characteristics, we predict the likelihood of these students being accepted as network partners in our experiment, had participants only seen their anonymized information (i.e., their experimental profiles). We call this likelihood "popularity score." In Table 2 below, we then relate the size and "quality" of students' networks (where quality is measured as the average popularity score of their friends) to their popularity score in bivariate regressions (Columns 1 and 2). Positive coefficients in these regressions would indicate that on average, the subjects' decisions in the experiment meaningfully relate to actual network formation in the months following

[26] In unreported regressions, we find that the coefficient on speaking a second language masks some differences by which language students speak (Leszczensky and Pink, 2019): While these analyses are noisy and not statistically significant, the point estimates suggest that there are homophilous preferences over speaking Turkish and somewhat homophilous preferences over speaking Arabian as a second language. In contrast, there seem to be common preferences favoring profiles with Spanish as second language.

the experiment, in the sense that more popular students (according to average choices in the experiment) are able to form larger actual networks and manage to form ties with more popular students. The table shows that both students' network size and quality are positively related with students' popularity score, although the estimate in the first column is insignificant. In Columns (3) and (4), we include field-specific fixed effects to account for the most important factors of a common environment of students and for student sorting on personality traits into fields of study (e.g., Buser et al., 2023; Adler et al., 2024). The table shows that both students' network size and network quality are now significantly related to students' own popularity scores.[27]

Table 2: Validation of Preference Estimates

	Network			
	Size	Quality	Size	Quality
	(1)	(2)	(3)	(4)
Pop. Score	0.51	0.29***	0.53*	0.27***
	(0.32)	(0.04)	(0.32)	(0.04)
Mean Dep. Var.	2.23	0.48	2.23	0.48
Obs.	691	388	691	388
Field FE	No	No	Yes	Yes

Note: This table shows regressions of how students' popularity scores (derived from average choice behavior in the experiment) relates to characteristics of the students' actual networks. As dependent variables, we use network size (actual number of friends) and network quality (average popularity score of actual friends). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Robustness checks in the appendix. In the Online Appendix, we explore the robustness of our estimates. First, we address the concern that students may perceive the experiment (and the networking event) as a way to meet potential dating partners (Fisman et al., 2006), or that they use gender as a proxy for characteristics not covered by the profile attributes (Gallen and Wasserman, forthcoming). To this end, we left out the gender dimension on the last screen showing fictitious profiles of potential network partners. In all other dimensions, the profiles were the same as on the first two screens. Figure A10 in the Appendix uses only the data from profiles without information on

[27]To illustrate the relevance of network size at university for students, in Table A3 in the Online Appendix we show that network size at university as measured in our data significantly predicts higher life satisfaction, lower depression and loneliness scores, and higher satisfaction with students' current networks and network sizes.

gender. The figure shows that our estimates remain qualitatively similar, although the estimates stem from a smaller sample and are noisier. If anything, homophilous preferences play even less of a role. In addition, Table A4 in the Online Appendix shows that our estimates by gender are qualitatively identical irrespective of the gender of the fictitious profile, again mitigating concerns that students view the experiment as a dating opportunity.[28] Second, we assess whether using only data from students interested in the networking event leads to different conclusions. Figure A11 in the Online Appendix shows that all our main findings are qualitatively identical.[29] Third, acknowledging that subjects not paying attention to the experimental instructions or the profile attributes could lead to attenuation of estimated parameters, Online Appendix Figure A12 shows that our results are robust to excluding data from respondents who did not spend a minimum amount of time on the choice screens.[30] Finally, Online Appendix Figure A13 shows that coding *maybe* choices as indicating acceptance of a profile rather than rejection does not alter any of our main conclusions.

"Sorting" into network connections on preferences. To assess the wider role of preferences for the formation of network connections in our sample, we test whether students "sort" into connections in the sense that students who have friends with specific characteristics value these characteristics more than students who do not have such friends.[31] To this end, we estimate students' preferences for specific characteristics in the experimental data by whether at least one friend they named in the follow-up survey has these characteristics. Figure A14 in the Online Appendix shows that students have similar preferences for the characteristics of their potential network partners irrespective of the prevalence of characteristics in their own networks. This is not in line with students sorting into friendships based on their preferences for specific characteristics.

[28]We also find that neither female nor male students value ambition (stated study motive to have a successful career) differentially by profile gender, again speaking against students perceiving contacts as potential dating partners (Bursztyn et al., 2017).

[29]Figure A11 also eases worries that participants in the experiment consider correlated unobservable characteristics of fictitious profiles instead of the variables we display. The reason is that, if that was true, we would expect students who intend to participate in the networking event to have different preferences than those who do not and who consequently can follow the instruction to assume identical unobservable characteristics of fictitious profiles. Instead, we find that preferences are similar irrespective of students' intent to join the networking event.

[30]Figure A12 excludes the fastest 10% of subjects (i.e., those who spent less than 13 seconds on a choice screen on average).

[31]We call this "sorting" in analogy to labor market sorting where workers tend to self-select into jobs with specific non-wage amenities based on their preferences for these (see, e.g., Rosen, 1986; Maestas et al., 2023; Nagler et al., forthcoming).

5 Conclusion

Homophily is a common feature of social networks, but there is little research to what extent it is due to homophilous preferences. As a result, little is known about whether the strong homophily in most people's social network is something people value on average. Alternatively, people could actually prefer less homophilous networks, but barriers to form social ties with network partners of their preferred type could prevent them from actually building such networks. Apart from self-selection into environments where social networks are formed, barriers to connect with preferred network partners could be due to common preferences (i.e., a situation where all subjects want to connect with peers carrying a specific characteristic, irrespective of their own type).

In this paper, we combine survey, administrative, and experimental data to investigate whether homophilous preferences are a likely driver of homophily in a social network of university students. Studying the actual social network, we demonstrate homophily in most of the characteristics and personality traits in our data. While we extend the literature by analyzing homophily in dimensions of personality that have not been studied before, we confirm the finding that homophily is a very wide-spread phenomenon characterizing social networks.

We then leverage our stated-choice experiment to show that the observed patterns of homophily are unlikely, however, to be explained by homophilous preferences. Rather, our results imply common preferences for certain traits and most characteristics, especially altruism and conscientiousness. The few dimensions where we find at least suggestive evidence of homophilous preferences include competitiveness and study fields, especially the humanities. Overall, we conclude that homophilous preferences are the exception rather than the norm.

References

Abele, A. E. and S. Brack (2013): "Preference for other persons' traits is dependent on the kind of social relationship," *Social Psychology*, 44, 84–89.

Abele, A. E., N. Ellemers, S. T. Fiske, A. Koch, and V. Yzerbyt (2020): "Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups." *Psychological Review*, 128, 290–314.

Adler, K., F. Kosse, M. Nagler, and J. Rincke (2024): "Students' Selection, Matching (and Adaption): The Role of Personality and Preferences," *Mimeo*.

Alan, S., C. Baysan, M. Gumren, and E. Kubilay (2021): "Building social cohesion in ethnically mixed schools: An intervention on perspective taking," *Quarterly Journal of Economics*, 136, 2147–2194.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong (2018): "Social Connectedness: Measurement, Determinants, and Effects," *Journal of Economic Perspectives*, 32, 259–280.

Bailey, M., D. Johnston, M. Koenen, T. Kuchler, D. Russel, and J. Stroebel (2024): "Social Networks Shape Beliefs and Behavior: Evidence from Social Distancing during the COVID-19 Pandemic," *Journal of Political Economy Microeconomics*, forthcoming.

Beaman, L., N. Keleher, and J. Magruder (2018): "Do job networks disadvantage women? Evidence from a recruitment experiment in Malawi," *Journal of Labor Economics*, 36, 121–157.

Beierlein, C., A. Kovaleva, C. J. Kemper, and B. Rammstedt (2012): "Ein Messinstrument zur Erfassung subjektiver Kompetenzerwartungen: Allgemeine Selbstwirksamkeit Kurzskala (ASKU)," GESIS Working Papers, GESIS – Leibniz-Institut für Sozialwissenschaften.

Bhargava, P., D. Chen, M. Sutter, and C. Terrier (2022): "Homophily and transmission of behavioral traits in social networks," IZA Discussion Papers No. 15840.

Bramoullé, Y., H. Djebbari, and B. Fortin (2020): "Peer effects in networks: A survey," *Annual Review of Economics*, 12, 603–629.

Bursztyn, L., T. Fujiwara, and A. Pallais (2017): "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, 107, 3288–3319.

Buser, T., R. Ahlskog, M. Johannesson, and S. Oskarsson (2023): "Occupational sorting on genes," *Tinbergen Institute Discussion Paper 2022-06/I*.

Buser, T., M. Niederle, and H. Oosterbeek (forthcoming): "Can Competitiveness Predict Education and Labor Market Outcomes? Evidence from Incentivized Choice and Survey Measures," *Review of Economics and Statistics*.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013): "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 81, 855–882.

Chetty, R., M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, et al. (2022): "Social capital II: determinants of economic connectedness," *Nature*, 608, 122–134.

Currarini, S., M. O. Jackson, and P. Pin (2009): "An Economic Model of Friendship: Homophily, Minorities, and Segregation," *Econometrica*, 77, 1003–1045.

Elwert, F., T. Keller, and A. Kotsadam (2023): "Rearranging the Desk Chairs: A Large Randomized Field Experiment on the Effects of Close Contact on Interethnic Relations," *American Journal of Sociology*, 128, 1809–1840.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018): "Global Evidence on Economic Preferences," *Quarterly Journal of Economics*, 133, 1645–1692.

Falk, A., A. Becker, T. J. Dohmen, D. Huffman, and U. Sunde (2023): "The Preference Survey Module: A Validated Instrument for Meausuring Risk, Time, and Social Preferences," *Management Science*, 69, 1935–2545.

Feld, S. and B. Grofman (2011): "Homophily and the Focused Organization of Ties," in *The Oxford Handbook of Analytical Sociology*, ed. by P. Bearman and P. Hedström, Oxford University Press, 521–543.

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2006): "Gender differences in mate selection: Evidence from a speed dating experiment," *Quarterly Journal of Economics*, 121, 673–697.

Flashman, J. (2011): "Academic Achievement and Its Impact on Friend Dynamics," *Sociology of Education*, 85, 61–80.

Gallen, Y. and M. Wasserman (forthcoming): "Does Information Affect Homophily?" *Journal of Public Economics*.

Gerlitz, J.-Y. and J. Schupp (2005): "Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP," Diw research notes 2005-4, Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).

Girard, Y., F. Hett, and D. Schunk (2015): "How individual characteristics shape the structure of social networks," *Journal of Economic Behavior & Organization*, 115, 197–216.

Golsteyn, B. H. H., A. Non, and U. Zölitz (2021): "The impact of peer personality on academic achievement," *Journal of Political Economy*, 129, 1052–1099.

Griffith, A. (2022): "Name your friends, but only five? The importance of censoring in peer effects estimates using social network data," *Journal of Labor Economics*, 40, 779–805.

Hitsch, G. J., A. Hortaçsu, and D. Ariely (2010): "Matching and sorting in online dating," *American Economic Review*, 100, 130–163.

Jackson, M. O. (2010): *Social and economic networks*, Princeton University Press, Princeton.

——— (2014): "Networks in the Understanding of Economic Behaviors," *Journal of Economic Perspectives*, 28, 3–22.

——— (2021): "Inequality's economic and social roots: The role of social networks and homophily," Mimeo.

Jackson, M. O., S. M. Nei, E. Snowberg, and L. Yariv (2023): "The dynamics of networks and homophily," NBER Working Paper No. 30815.

Jackson, M. O., B. W. Rogers, and Y. Zenou (2017): "The economic consequences of social-network structure," *Journal of Economic Literature*, 55, 49–95.

Knecht, A., T. A. B. Snijders, C. Baerveldt, C. E. G. Steglich, and W. Raub (2010): "Friendship and Delinquency: Selection and Influence Processes in Early Adolescence," *Social Development*, 19, 494–514.

Kossinets, G. and D. J. Watts (2009): "Origins of Homophily in an Evolving Social Network," *American Journal of Sociology*, 115, 405–450.

Kruse, H., S. Smith, F. van Tubergen, and I. Maas (2016): "From neighbors to school friends? How adolescents' place of residence relates to same-ethnic school friendships," *Social Networks*, 44, 130–142.

Leszczensky, L. and S. Pink (2019): "What Drives Ethnic Homophily? A Relational Approach on How Ethnic Identification Moderates Preferences for Same-Ethnic Friends," *American Sociological Review*, 84, 394–419.

Lomi, A., T. A. Snijders, C. E. Steglich, and V. J. Torló (2011): "Why are some more peer than others? Evidence from a longitudinal study of social networks and individual academic performance," *Social Science Research*, 40, 1506–1520.

Maestas, N., K. J. Mullen, D. Powell, T. von Wachter, and J. D. Wenger (2023): "The Value of Working Conditions in the United States and Implications for the Structure of Wages," *American Economic Review*, 113, 2007–2047.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001): "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27, 415–444.

Melamed, D., M. Sweitzer, B. Simpson, J. Z. Abernathy, A. Harrell, and C. W. Munn (2020): "Homophily and Segregation in Cooperative Networks," *American Journal of Sociology*, 125, 1084–1127.

Nagler, M., J. Rincke, and E. Winkler (forthcoming): "High-Pressure, High-Paying Jobs?" *The Review of Economics and Statistics*.

Neyt, B., S. Vandenbulcke, and S. Baert (2019): "Are men intimidated by highly educated women? Undercover on Tinder," *Economics of Education Review*, 73, 101914.

Niederle, M. and L. Vesterlund (2007): "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 122, 1067–1101.

Norris, J. (2020): "Peers, parents, and attitudes about school," *Journal of Human Capital*, 14, 290–342.

Ollroge, R. and T. Sawert (2022): "The Cultural Dimension of the Globalization Divide. Do Lifestyle Signals affect Cosmopolitans' Willingness to interact?" *Zeitschrift für Soziologie*, 51, 263–277.

Ong, D. (2016): "Education and income attraction: An online dating field experiment," *Applied Economics*, 48, 1816–1830.

Porter, C. M. and J. R. Rigby (2019): "Relationship context and personality shape people's preferences for network relationship partners," *Personal Relationships*, 26, 310–330.

Rao, G. (2019): "Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools," *American Economic Review*, 109, 774–809.

Rosen, S. (1986): "The Theory of Equalizing Differences," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter and R. Layard, North Holland, vol. 1, chap. 12, 641–692.

Sacerdote, B. (2014): "Experimental and quasi-experimental analysis of peer effects: Two steps forward?" *Annual Review of Economics*, 6, 253–272.

Sawka, K. J., G. R. McCormack, A. Nettel-Aguirre, P. Hawe, and P. K. Doyle-Baker (2013): "Friendship networks and physical activity and sedentary behavior among youth: a systematized review," *International Journal of Behavioral Nutrition and Physical Activity*, 10, 130.

Schaefer, D. R., O. Kornienko, and A. M. Fox (2011): "Misery Does Not Love Company: Network Selection Mechanisms and Depression Homophily," *American Sociological Review*, 76, 764–785.

Schaefer, D. R. and D. A. Kreager (2020): "New on the Block: Analyzing Network Selection Trajectories in a Prison Treatment Program," *American Sociological Review*, 85, 709–737.

Shan, X. and U. Zölitz (2022): "Peers affect personality development," CEPR Discussion Paper No. DP17241.

Shkoza, L., D. Uysal, and W. Pohlmeier (2020): "Heterogeneity in Network Peer Effects," *Mimeo*.

Skopek, J., F. Schulz, and H.-P. Blossfeld (2011): "Who Contacts Whom? Educational Homophily in Online Mate Selection," *European Sociological Review*, 27, 180–195.

Small, M. L. and D. Pager (2020): "Sociological Perspectives on Racial Discrimination," *Journal of Economic Perspectives*, 34, 49–67.

Smith, J. A., M. McPherson, and L. Smith-Lovin (2014a): "Social Distance in the United States: Sex, Race, Religion, Age, and Education Homophily among Confidants, 1985 to 2004," *American Sociological Review*, 79, 432–456.

Smith, S., I. Maas, and F. van Tubergen (2014b): "Ethnic ingroup friendships in schools: Testing the by-product hypothesis in England, Germany, the Netherlands and Sweden," *Social Networks*, 39, 33–45.

Snijders, T. A. and A. Lomi (2019): "Beyond homophily: Incorporating actor variables in statistical network models," *Network Science*, 7, 1–19.

Tekles, A., K. Auspurg, and L. Bornmann (2022): "Same-gender citations do not indicate a substantial gender homophily bias," *PLoS ONE*, 17, e0274810.

Weber, H., M. Schwenzer, and S. Hillmert (2020): "Homophily in the formation and development of learning networks among university students," *Network Science*, 8, 469–491.

Wimmer, A. and K. Lewis (2010): "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook," *American Journal of Sociology*, 116, 583–642.

Zeltzer, D. (2020): "Gender homophily in referral networks: Consequences for the medicare physician earnings gap," *American Economic Journal: Applied Economics*, 12, 169–197.

Zuckerman, D. (2024): "Multidimensional homophily," *Journal of Economic Behavior & Organization*, 218, 486–513.

A Appendix

A.1 Additional Summary Statistics

Table A1: Sorting into the Experiment

	Full sample	Participants	Non-participants	Difference (3)-(2)
	(1)	(2)	(3)	(4)
Demographic characteristics				
Female	0.57	0.57	0.57	-0.00
	(0.50)	(0.50)	(0.50)	(0.02)
Age	19.44	19.40	19.46	0.06
	(2.21)	(1.99)	(2.35)	(0.07)
Secondary GPA	2.75	2.86	2.68	-0.18***
	(0.68)	(0.68)	(0.68)	(0.02)
Field of study				
Teaching	0.15	0.15	0.15	-0.00
	(0.35)	(0.35)	(0.35)	(0.01)
Humanities	0.12	0.08	0.15	0.07***
	(0.32)	(0.27)	(0.35)	(0.01)
Engenieering	0.16	0.15	0.17	0.02*
	(0.37)	(0.36)	(0.38)	(0.01)
Natural sciences	0.15	0.14	0.15	0.02
	(0.35)	(0.34)	(0.36)	(0.01)
Law	0.09	0.07	0.10	0.03***
	(0.28)	(0.25)	(0.30)	(0.01)
Economics/Business	0.18	0.23	0.15	-0.08***
	(0.39)	(0.42)	(0.36)	(0.01)
Medicine	0.08	0.09	0.06	-0.03***
	(0.26)	(0.29)	(0.24)	(0.01)
Social sciences	0.08	0.10	0.06	-0.04***
	(0.27)	(0.30)	(0.24)	(0.01)
N	3,572	1,491	2,081	3,572

Note: This table provides summary statistics describing sorting into the sample of students who participated in the stated-choice experiment. Column (1) shows statistics for the full cohort of students belonging to the 2023 cohort, all of whom were invited to the experiment. Columns (2) to (4) differentiate between participants and non-participants. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

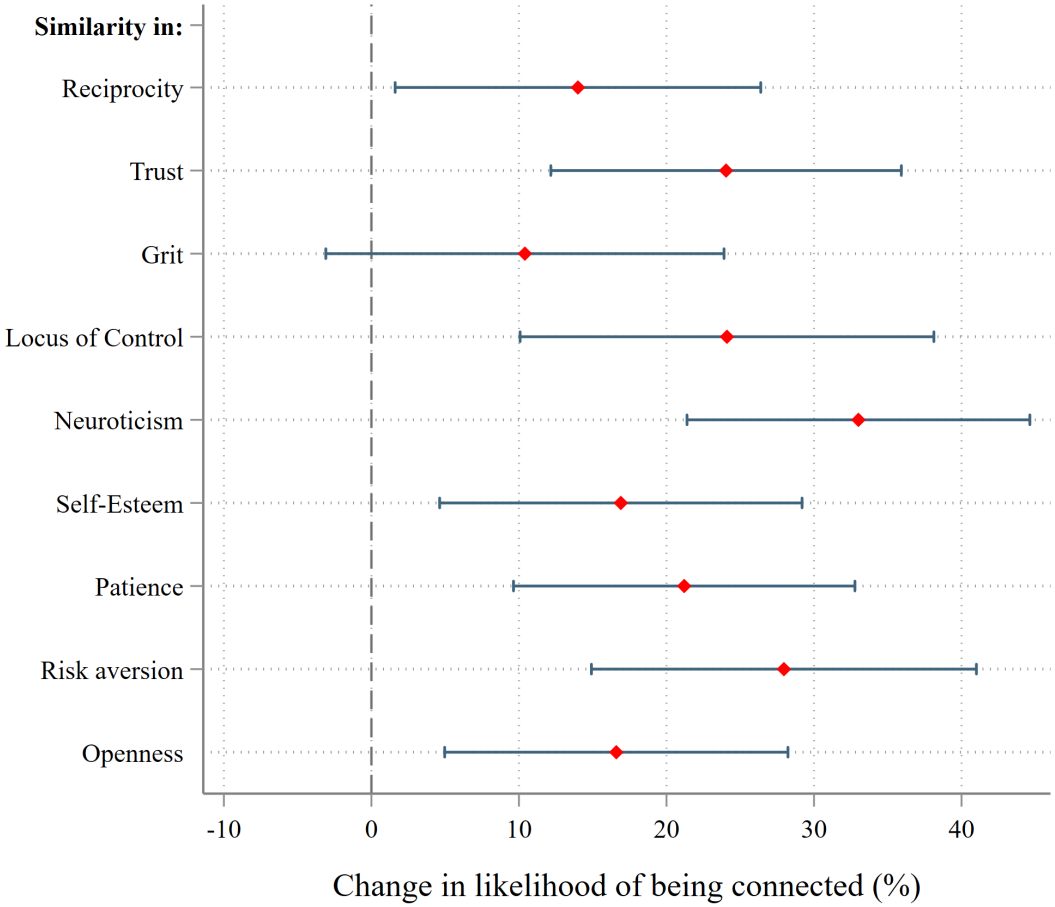
Table A2: Participation: Follow-up Survey on Existing Networks

	Participants	Follow-up	Difference (2)-(1)	Students with data on network	Students without data on network	Difference (5)-(4)
	(1)	(2)	(3)	(4)	(5)	(6)
Demographic characteristics						
Female	0.57	0.57	-0.00	0.56	0.60	0.03
	(0.50)	(0.50)	(0.03)	(0.50)	(0.49)	(0.05)
Age	19.40	19.36	0.06	19.23	19.72	0.49**
	(1.99)	(2.26)	(0.11)	(1.77)	(3.22)	(0.21)
Secondary GPA	2.86	2.94	-0.13***	2.95	2.90	-0.05
	(0.68)	(0.69)	(0.04)	(0.68)	(0.71)	(0.06)
Field of study						
Teaching	0.15	0.16	-0.03	0.15	0.21	0.06*
	(0.35)	(0.37)	(0.02)	(0.36)	(0.41)	(0.03)
Humanities	0.08	0.08	-0.00	0.07	0.08	0.01
	(0.27)	(0.27)	(0.01)	(0.26)	(0.28)	(0.03)
Engenieering	0.15	0.15	-0.01	0.15	0.16	0.01
	(0.36)	(0.36)	(0.02)	(0.36)	(0.37)	(0.03)
Natural sciences	0.14	0.14	-0.00	0.14	0.13	-0.01
	(0.34)	(0.35)	(0.02)	(0.35)	(0.34)	(0.03)
Law	0.07	0.06	0.01	0.06	0.06	-0.00
	(0.25)	(0.24)	(0.01)	(0.24)	(0.24)	(0.02)
Economics/Business	0.23	0.18	0.07***	0.20	0.14	-0.05
	(0.42)	(0.39)	(0.02)	(0.40)	(0.35)	(0.04)
Medicine	0.09	0.10	-0.01	0.09	0.12	0.03
	(0.29)	(0.30)	(0.02)	(0.29)	(0.33)	(0.03)
Social sciences	0.10	0.12	-0.02	0.13	0.09	-0.04
	(0.30)	(0.32)	(0.02)	(0.33)	(0.29)	(0.03)
N	1,491	570	1,491	416	154	570

Note: This table provides summary statistics for the sample of students participating in the follow-up survey on existing networks. Columns (1) to (3) show statistics for participants and non-participants in the follow-up survey in july 2024. Columns (4) to (6) report statistics for participants in the follow-up survey, differentiating between students who provided data on their social network and students who did not. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.2 More Information on Homophily in Observable Characteristics

Figure A1: Homophily in Personality Characteristics



Note: This figure is based on observational data and shows homophily in additional traits and preferences. Each coefficient measures homophily in the respective dimension and corresponds to a separate regression based on equation (2). The sample consists of a subset of 384 participants and 443 subjects named as friends, where we have complete information on all traits and preferences in the cohort survey. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

A.3 Observed Homophily in Other Student Cohorts

Apart from our main sample, we collected data on the social networks of students of the cohorts 2020 and 2021, that participated in the respective cohort surveys[32], at the conclusion of their studies. This allows us to describe homophily within networks that were not recently formed, but rather developed over three years at university.

Cohort of 2020. In July 2023, all students who had taken part in the 2020 cohort survey were invited to participate in a follow-up survey, with 1,300 students responding. In the follow-up survey, we again asked participants to name their closest friends at the university. Of the 1,300 respondents, 602 students provided information on at least one friend that we could identify through administrative data. 1,503 individuals from the 2020 cohort were mentioned as a friend at least once, with certain students being named multiple times. Overall, we are able to describe 1,834 binary friendship relationships for the cohort of 2020.

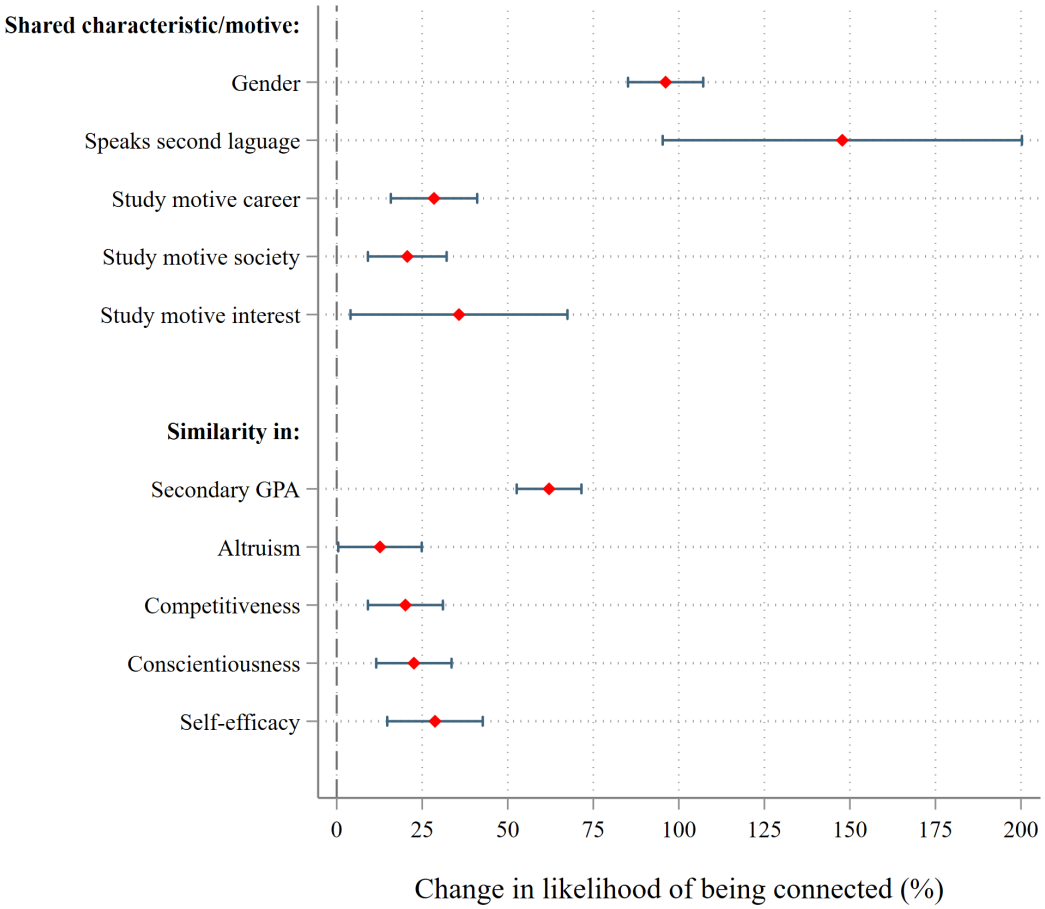
Again, information on characteristics and traits is fully available for all 602 participants in the follow-up survey who shared their network data. In total, our dataset includes personality traits for 559 of the 1,503 students named as friends. We can describe 750 friendship connections based on characteristics and traits.

Cohort of 2021. Students from the 2021 cohort survey were invited to join a follow-up survey in July 2024. Of the 854 respondents, 563 students provided data on their closest friends at university, which we were able to identify. In total, 1,493 students were mentioned as friends, resulting in data for 1,873 unique friendship connections.

For 452 students who were named as friends, data on personality traits is available. Altogether, we can describe 660 friendship connections in terms of both characteristics and traits.

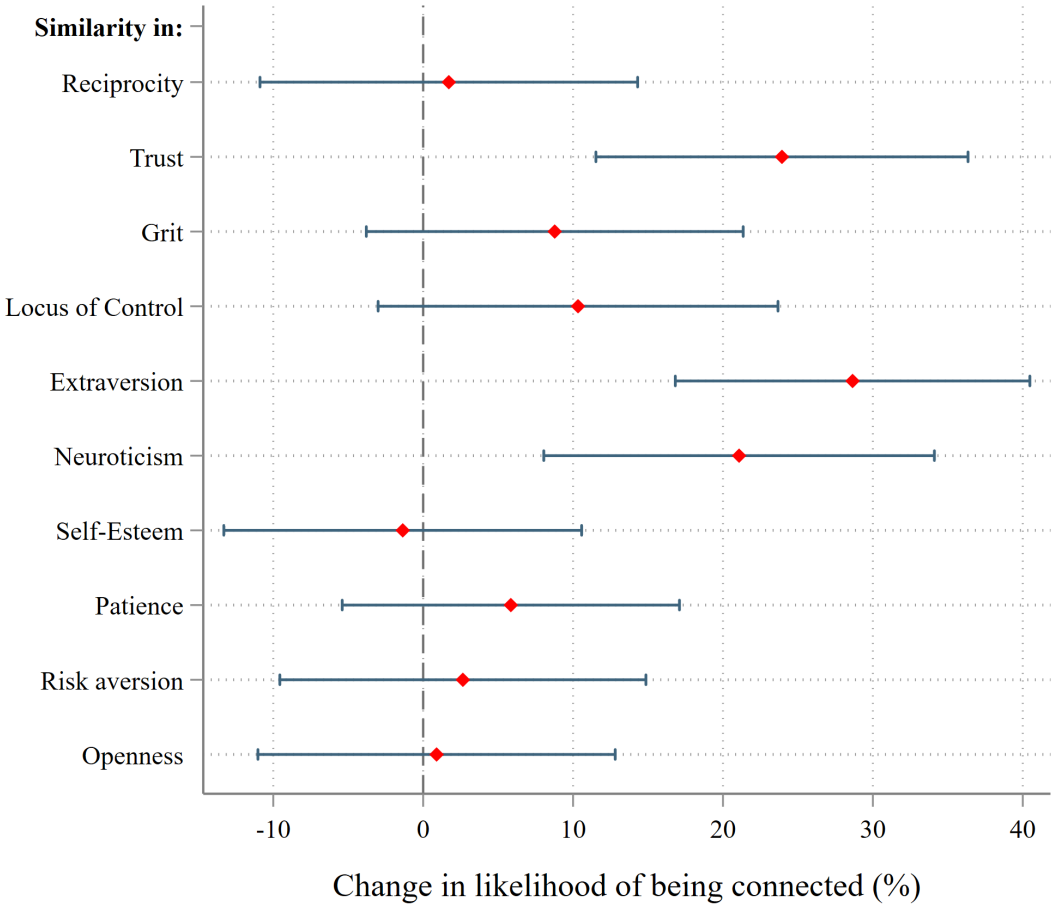
[32]For details on the cohort surveys refer to Section 2 or (Adler et al., 2024).

Figure A2: Observed Homophily in Cohort 2020



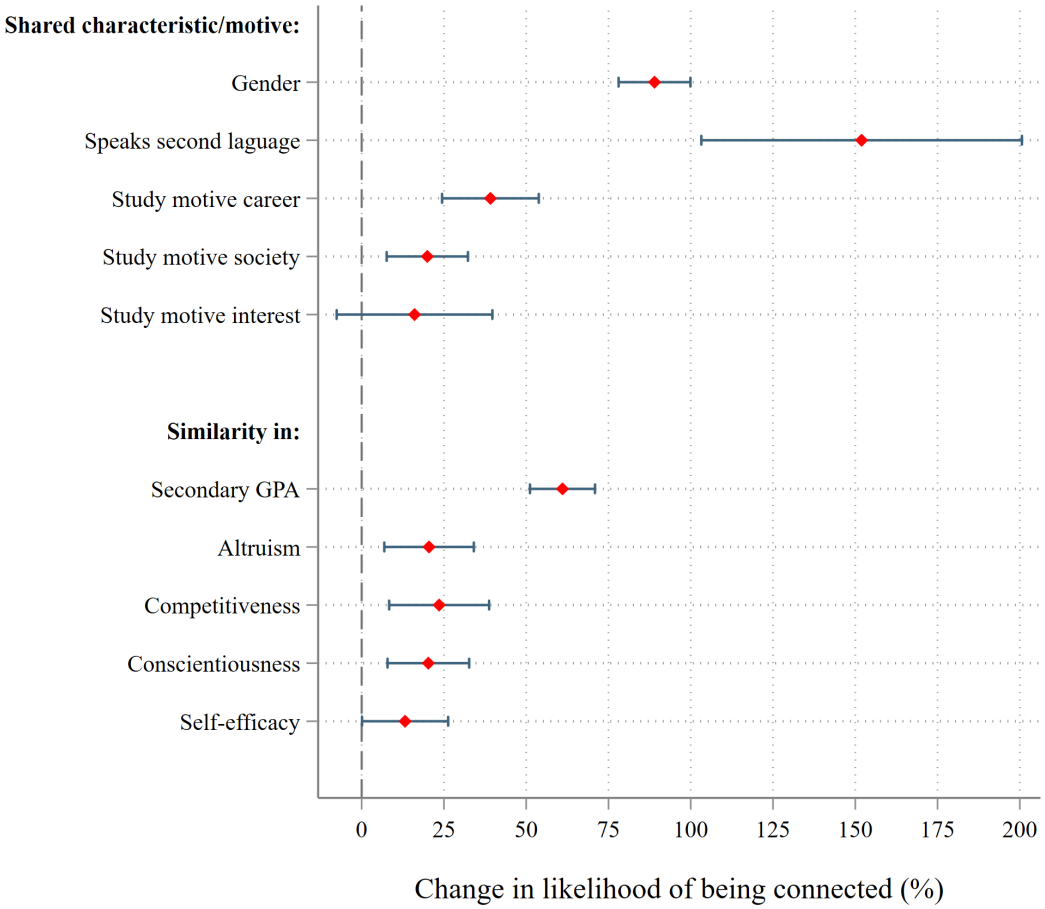
Note: This figure is based on observational data and shows homophily in demographic characteristics, study motives. and traits. Each coefficient measures homophily in the respective dimension and corresponds to a separate regression based on equation (1) or (2), depending on whether or not the respective dimension is binary. For gender and GPA, the sample consists of all (potential) connections between the 602 subjects naming at least one contact and the 1,503 subjects named at least once as friends. For all other dimensions, the sample consists of all (potential) connections between individuals and the 559 subjects named as friends who took part in the initial cohort survey in 2020. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

Figure A3: Homophily in Personality Characteristics in Cohort 2020



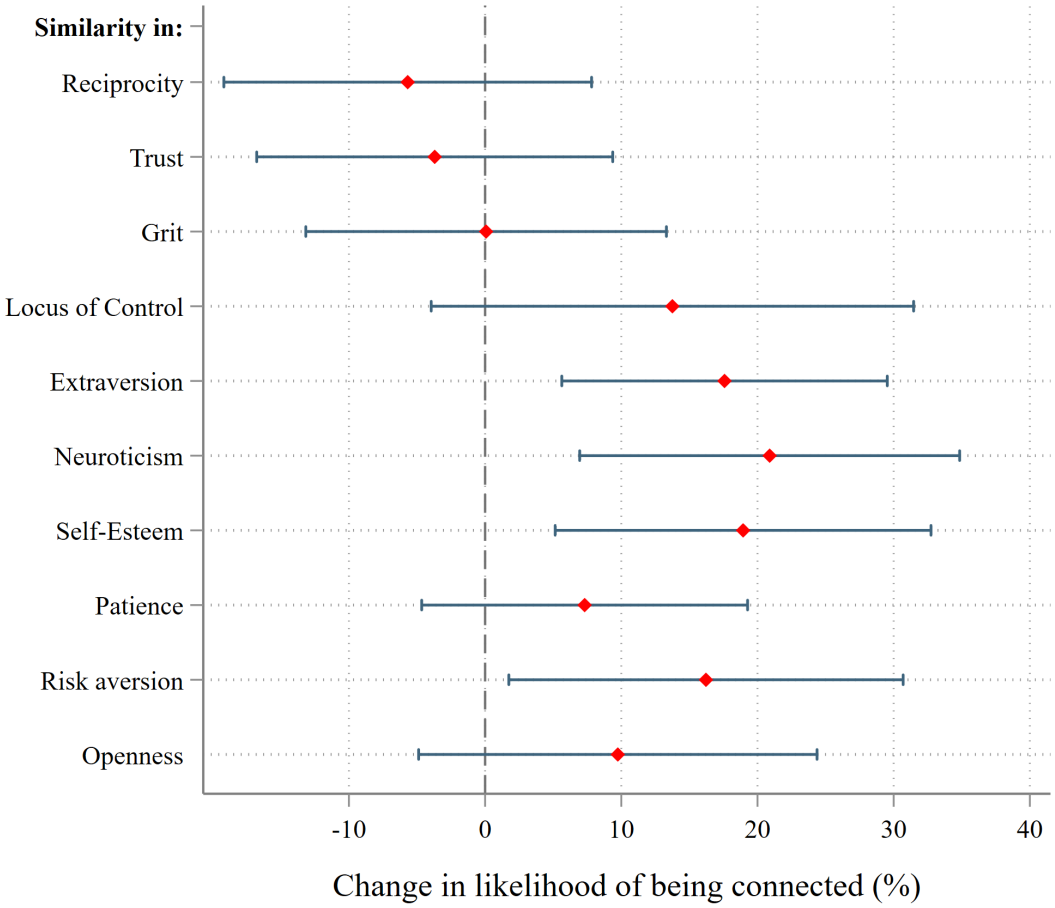
Note: This figure is based on observational data and shows homophily in additional traits and preferences for the cohort of 2020. Each coefficient measures homophily in the respective dimension and corresponds to a separate regression based on equation (2). The sample consists of a subset of 549 participants and 503 subjects named as friends, where we have complete information on all traits and preferences in the cohort survey. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

Figure A4: Observed Homophily in Cohort 2021



Note: This figure is based on observational data and shows homophily in demographic characteristics, study motives. and traits. Each coefficient measures homophily in the respective dimension and corresponds to a separate regression based on equation (1) or (2), depending on whether or not the respective dimension is binary. For gender and GPA, the sample consists of all (potential) connections between the 563 subjects naming at least one contact and the 1,493 subjects named at least once as friends. For all other dimensions, the sample consists of all (potential) connections between individuals and the 452 subjects named as friends who took part in the initial cohort survey in 2021. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

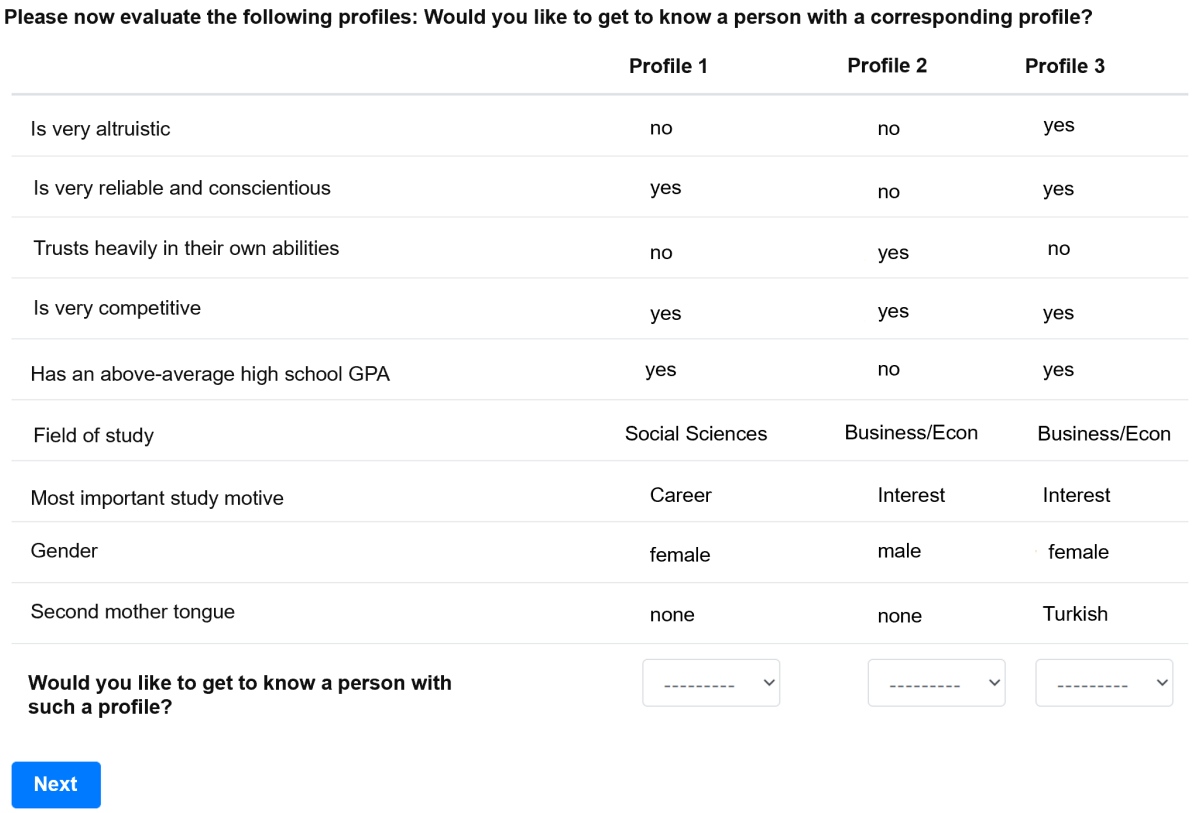
Figure A5: Homophily in Personality Characteristics in Cohort 2021



Note: This figure is based on observational data and shows homophily in additional traits and preferences for the cohort of 2021. Each coefficient measures homophily in the respective dimension and corresponds to a separate regression based on equation (2). The sample consists of a subset of 511 participants and 424 subjects named as friends, where we have complete information on all traits and preferences in the cohort survey. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

A.4 More Information on Stated-Choice Experiment

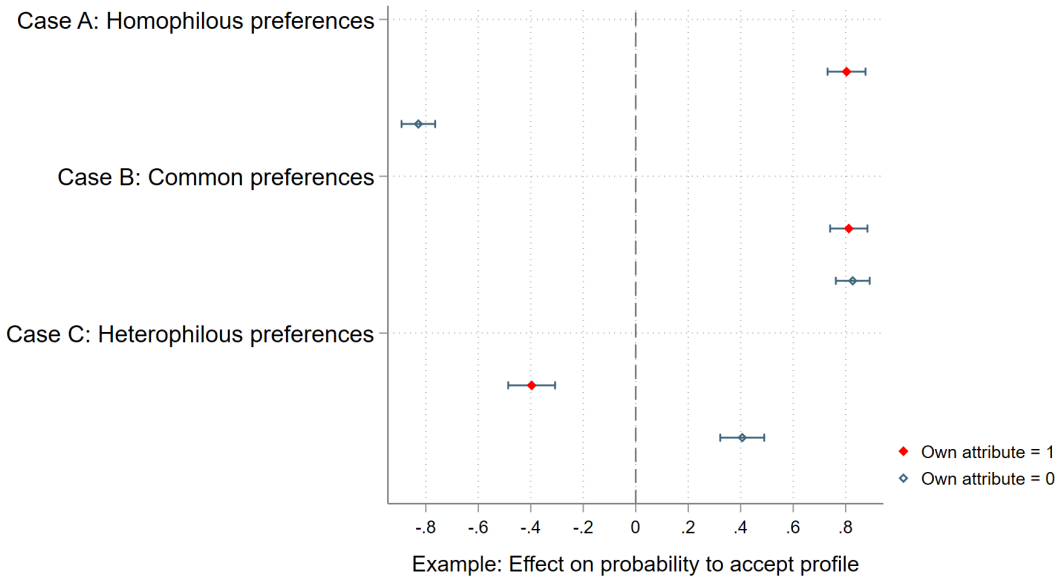
Figure A6: Sample Choice Screen



Note: This figure provides an example of a choice screen shown in the experiment. Individuals evaluated a total of six profiles (shown on two consecutive screens) with a complete set of attributes, and three more profiles (one screen) with profiles that did not include gender as an attribute. The choice options were "yes", "no", and "maybe".

A.5 Interpreting Our Estimates

Figure A7: Interpreting Coefficients: Fictitious Scenarios

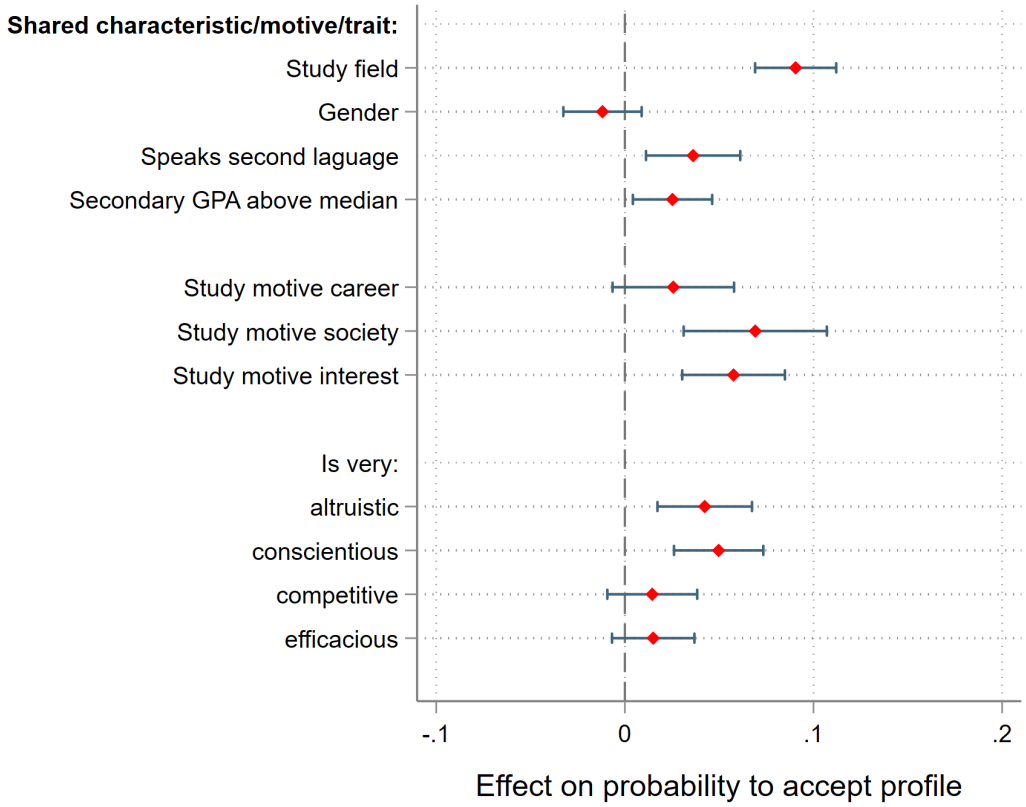


Note: This figure illustrates three scenarios based on equation (3). The red estimates indicate preferences of respondents that hold the respective attribute. Blue diamonds denote preferences of respondents who do not hold the respective attribute. In the first row the homophily scenario is shown. There is a positive preference for an attribute in a fictitious profile among respondents who have this attribute, and a negative preference when not. In the second row, the scenario of common preferences for an attribute is shown. Irrespective of whether respondents have a given attribute or not, the coefficient is positive, indicating a preference for this attribute. The third row shows the scenario of heterophilous preferences, where bearers of an attribute prefer non-bearers and vice-versa.

A.6 Naive Estimates and Average Common Components of Preferences

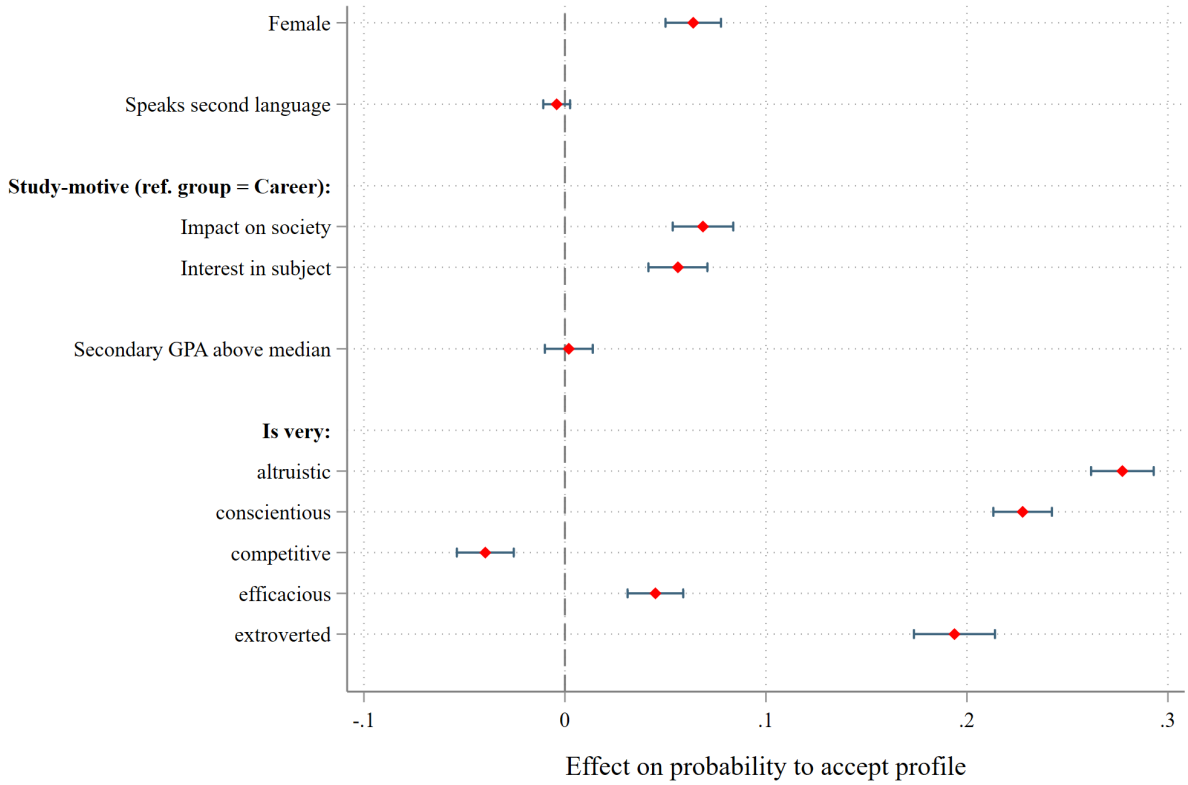
Given that for most of the characteristics and traits observed homophily cannot be explained by homophilous preferences, it is worth highlighting how important it is that our estimation approach based on equation (3) allows to identify homophilous, heterophilous, and common preferences in each dimension. To illustrate, this Appendix Section reports estimates resulting from two types of approaches that can produce misleading evidence regarding underlying preferences for network partners. The first approach is naive in the sense that it follows the logic of estimating homophily in an existing social network by conditioning only on shared characteristics, without taking into account own and profile characteristics separately. Figure A8 in this appendix shows that using this approach, we would estimate that preferences are "homophilous" in a number of dimensions, although Figure 2 shows that they are not. The second approach only conditions on profile characteristics in levels, without estimating separate effects by students' own characteristics (Appendix Figure A9). This approach produces estimates representing the average common component of preferences shown in Figure 2 and reveals that some profile characteristics are preferred relative to others: women, those with a study motive other than one's own career, and altruistic, conscientious and efficacious profiles are more likely to be accepted as potential network partners. This would not allow to estimate homophilous preferences, though.

Figure A8: Naive Estimates of Homophily in Preferences for Network Partners



Note: This figure shows estimated preferences for network partners by shared demographic characteristics, traits, and study motives without conditioning on own and profile attributes separately. The estimation equation is $accept_{ij} = c + \sum_{k=1}^{K} \rho^k \mathbb{1}[x_j^k = x_i^k] + \epsilon_{ij}$. Hence, coefficients represent a change in the probability of accepting a potential partner resulting from a shared attribute with the respective profile. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

Figure A9: Estimating Average Common Components of Preferences



Note: This figure shows estimated preferences for network partners by demographic characteristics, traits, and study motives without conditioning on own attribute values. The estimation equation is $accept_{ij} = c + \sum_{k=1}^{K} \rho^k(x_j^k = 1) + \epsilon_{ij}$. Hence, coefficients represent a change in the probability of accepting a potential partner resulting from a profile bearing the respective attribute. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

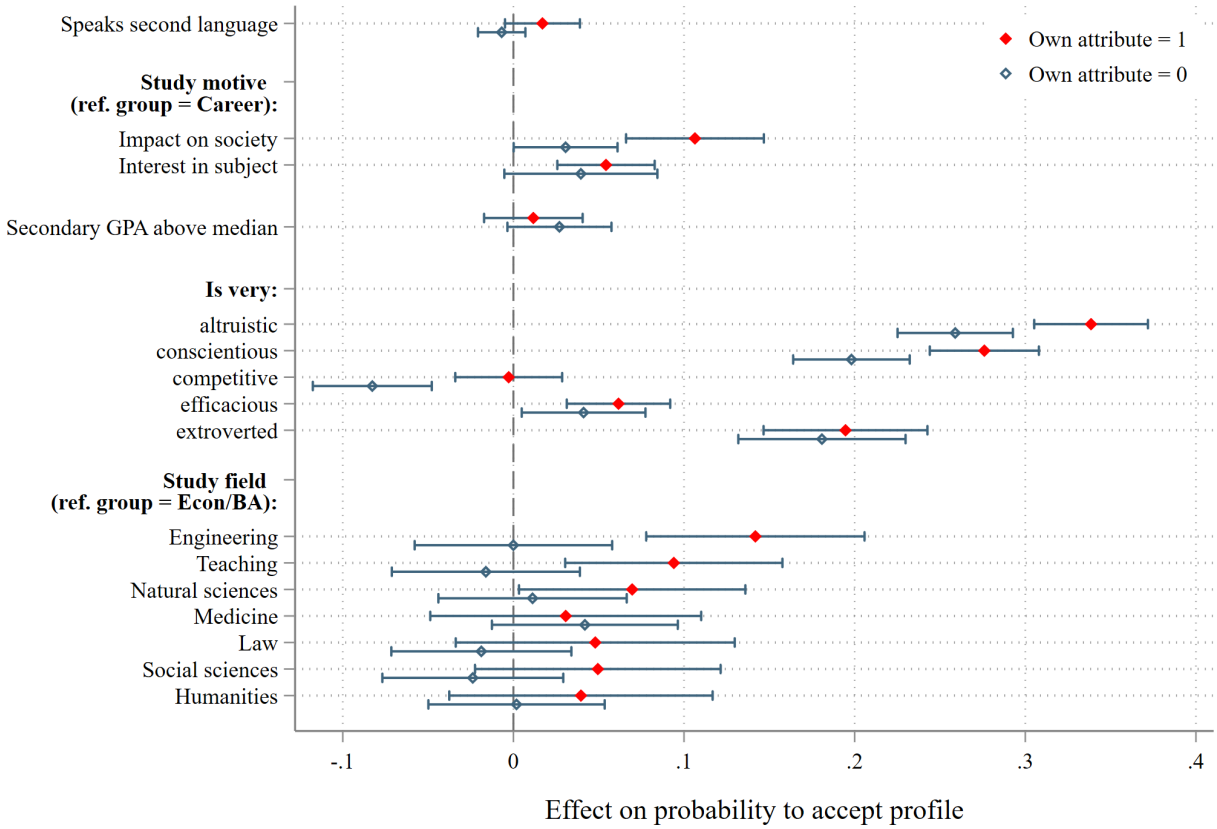
Table A3: Relevance of Larger Network Size at University

	Mental Health			Network			Information
	Life Sat.	Depr.	Lonelin.	Content	Want larger..		Index
	(standardized)				at Uni	Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
# Friends	0.05**	-0.04**	-0.07***	0.26***	-0.06	-0.06	0.04
	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.05)	(0.03)
Mean Dep. Var.	-0.00	-0.00	0.00	6.46	6.02	5.06	5.14
Obs.	687	687	679	689	690	689	648

This table shows results from regressions of different measures of mental health, network satisfaction, and informedness of students on students' stated network size at university. The number of friends is measured regardless of whether we can match the named friend using administrative data. Therefore, the number of observations is larger compared to the homophily analysis. To measure overall life satisfaction in column (1), we rely on the SOEP question: "How satisfied are you with your life, all things considered?" measured on an 11-point Likert scale. A higher score corresponds to higher life satisfaction. To measure depression symptoms, we rely on the PHQ-4 questionnaire in column (2), and to measure loneliness, we utilize the UCLA-3 questionnaire in column (3). The scores from the individual questions in each questionnaire are summed and standardized, resulting in a single score for both depression and loneliness. Higher scores indicate more severe mental health problems. In column (4), contentment is derived from the question: "Generally speaking, how satisfied are you with your social environment at the university?" measured on a 9-point Likert scale, with higher values indicating greater satisfaction. Additionally, we asked whether students would like to have more contact with other students in column (5) and with other people in general in column (6). Higher values indicate stronger agreement with the question, measured on a 9-point Likert scale. The information index in column (7) is based on five questions. We asked students how well informed they feel about news, events, and opportunities related to the university and their education (on a 9-point Likert scale). Higher values indicate better informedness. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.7 Robustness Analyses

Figure A10: Preference Estimates Excluding Gender



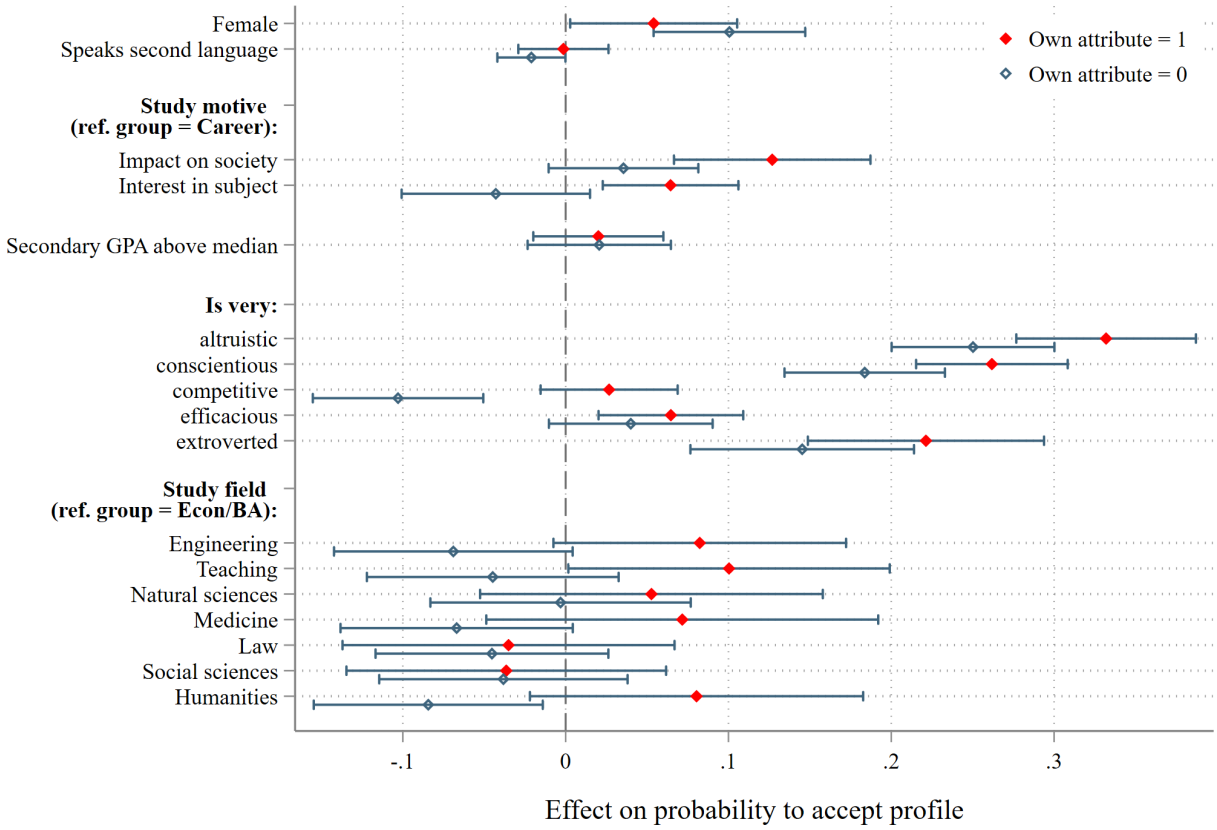
Note: This figure shows preferences for network partners for a set of choice experiments excluding gender from the profiles. Based on equation (3), coefficients represent changes in the probability of a profile being selected as a potential network partner. The red estimates indicate preferences of respondents that hold the respective attribute. Blue diamonds denote preferences of respondents who do not hold the respective attribute. The dependent variable is an indicator equal to 1 if respondent i accepts a profile j as a potential partner and 0 otherwise. The sample consists of all decisions made by the 1,494 participants when considering profiles excluding gender, and thus $1494 \times 3 = 4482$ observations. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

Table A4: Preferences for Network Partners by Subject and Profile Gender

	(1)	(2)	(3)	(4)
Subject's gender	Female	Female	Male	Male
Profile's gender	Male	Female	Female	Male
Above median GPA	-0.04	-0.06**	-0.00	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)
Self: Above median GPA	-0.04	-0.03	-0.03	-0.02
	(0.03)	(0.03)	(0.04)	(0.03)
Above median GPA × Self: Above median GPA	0.05	0.08**	0.03	0.03
	(0.04)	(0.04)	(0.04)	(0.04)
Speaks second mother tongue	-0.01	0.02	-0.08**	-0.08**
	(0.03)	(0.03)	(0.03)	(0.03)
Self: Speaks second mother tongue	-0.04	-0.05**	0.00	-0.00
	(0.03)	(0.03)	(0.03)	(0.03)
Speaks second mother tongue × Self: Speaks second mother tongue	0.07	0.05	0.11*	0.13**
	(0.05)	(0.05)	(0.06)	(0.06)
Altruistic	0.26***	0.30***	0.26***	0.23***
	(0.04)	(0.04)	(0.03)	(0.03)
Self: Altruistic	-0.00	0.00	-0.05	-0.05
	(0.03)	(0.03)	(0.04)	(0.04)
Altruistic × Self: Altruistic	0.08**	0.07	0.04	0.06
	(0.04)	(0.04)	(0.05)	(0.05)
Conscientious	0.24***	0.27***	0.24***	0.19***
	(0.03)	(0.03)	(0.04)	(0.04)
Self: Conscientious	-0.05*	-0.00	-0.06*	-0.06
	(0.03)	(0.03)	(0.04)	(0.04)
Conscientious × Self: Conscientious	0.08**	0.01	0.08*	0.12***
	(0.04)	(0.04)	(0.05)	(0.04)
Competitive	-0.11***	-0.09***	0.00	-0.04
	(0.03)	(0.03)	(0.05)	(0.05)
Self: Competitive	-0.06*	-0.02	-0.03	-0.04
	(0.03)	(0.03)	(0.04)	(0.04)
Competitive × Self: Competitive	0.11***	0.03	0.07	0.10*
	(0.04)	(0.04)	(0.06)	(0.05)
Efficacious	0.01	0.05*	0.09**	0.07**
	(0.03)	(0.03)	(0.04)	(0.04)
Self: Efficacious	-0.07**	-0.01	-0.03	-0.00
	(0.03)	(0.03)	(0.04)	(0.04)
Efficacious × Self: Efficacious	0.10**	0.01	0.03	0.01
	(0.04)	(0.04)	(0.05)	(0.04)
Study motive society	0.02	0.03	-0.04	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)
Self: Study motive society	-0.04	-0.03	0.02	0.01
	(0.03)	(0.03)	(0.04)	(0.03)
Study motive society × Self: Study motive society	0.01	0.04	0.03	0.07
	(0.04)	(0.04)	(0.06)	(0.06)
Study motive career	-0.08***	-0.09***	-0.13***	-0.11***
	(0.03)	(0.03)	(0.04)	(0.03)
Self: Study motive career	0.02	-0.00	-0.02	-0.06*
	(0.03)	(0.03)	(0.03)	(0.03)
Study motive career × Self: Study motive career	0.09**	0.08*	0.07	0.16***
	(0.04)	(0.04)	(0.05)	(0.05)
Observations	2540	2578	1912	1934

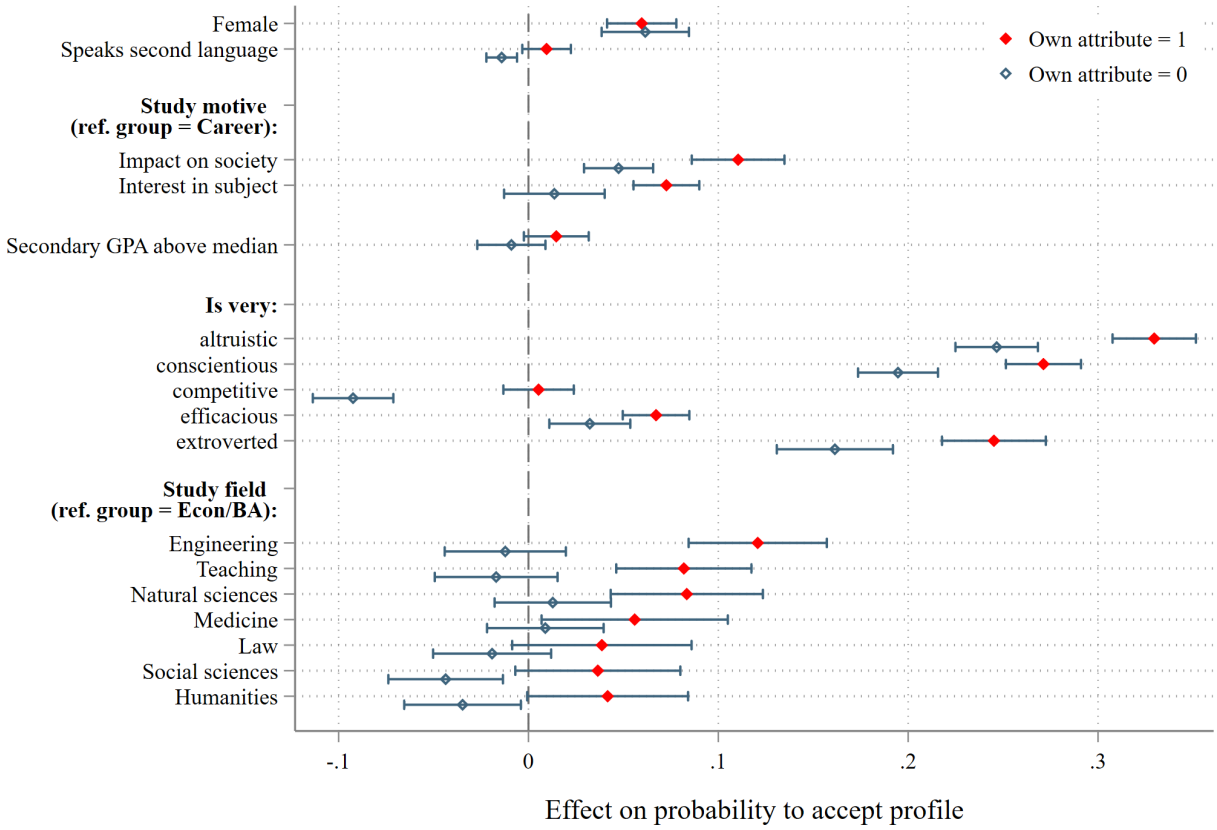
Note: This table shows results from our main specification for homophilous preferences (Equation 3) by subject and profile gender. Column (1) presents results for female students evaluating male profiles, Column (2) shows results for males evaluating females, Column (3) displays results for females evaluating females, and Column (4) presents results for males evaluating male profiles. Standard errors in parentheses allow for clustering at the student level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A11: Preference Estimates for Students Participating in Networking Event



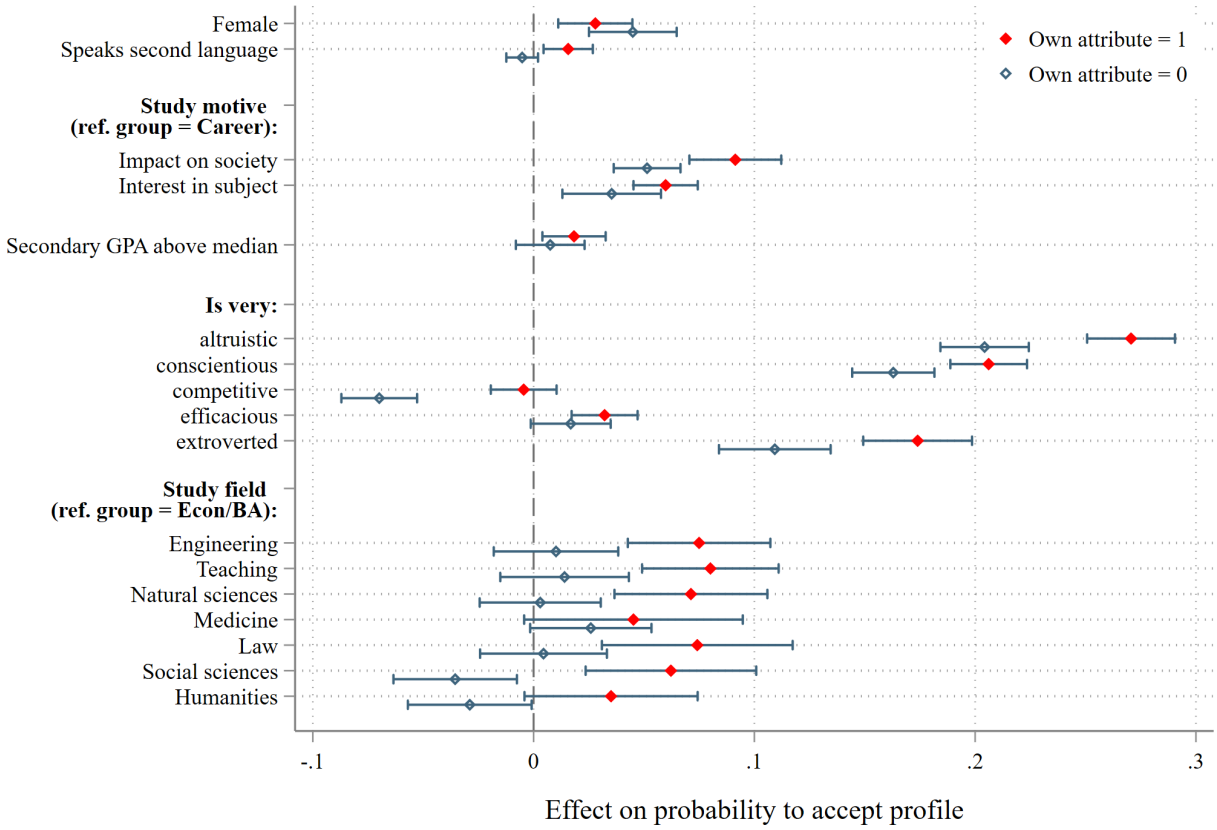
Note: This figure shows preferences for individuals who stated that they would be interested to participate in the online networking event. Based on equation (3), coefficients represent changes in the probability of a profile being selected as a potential network partner. The red estimates indicate preferences of respondents that hold the respective attribute. Blue diamonds denote preferences of respondents who do not hold the respective attribute. The dependent variable is an indicator equal to 1 if respondent i accepts a profile j as a potential partner and 0 otherwise. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

Figure A12: Preference Estimates Excluding Fastest Participants



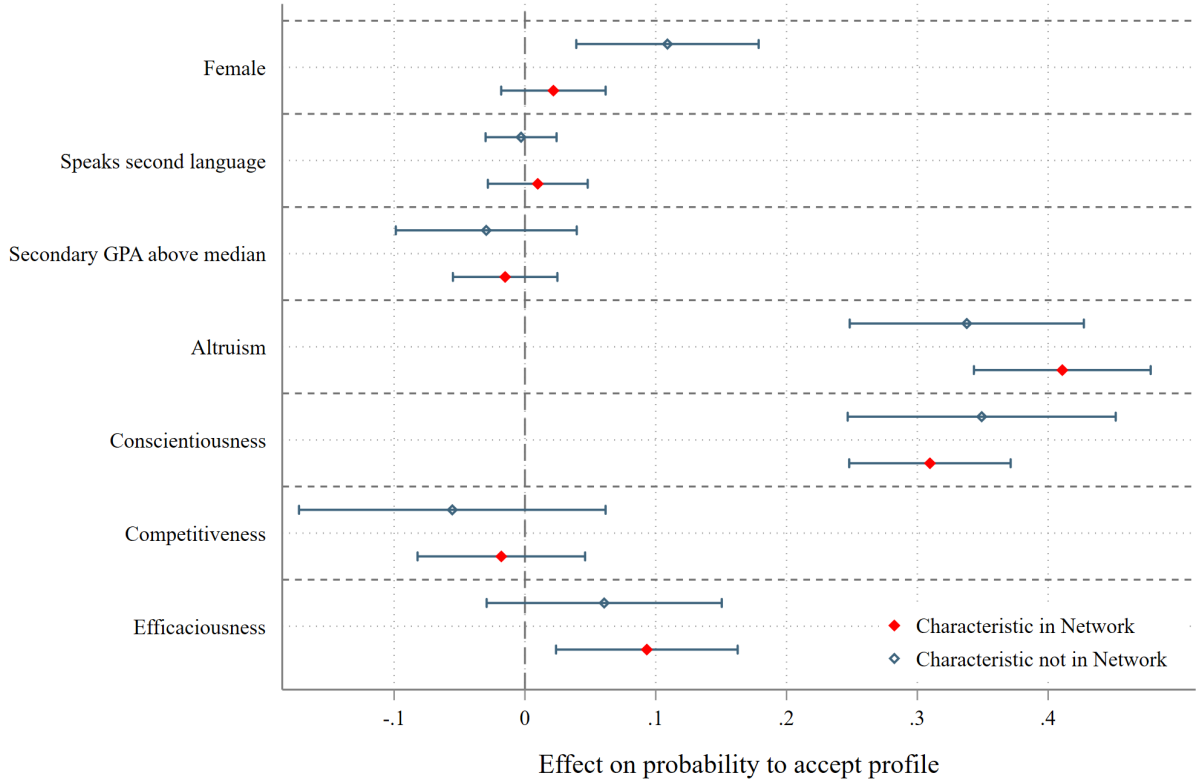
Note: This figure shows preferences for network partners, excluding the fastest 10% of individuals. Based on equation (3), coefficients represent changes in the probability of a profile being selected as a potential network partner. The red estimates indicate preferences of respondents that hold the respective attribute. Blue diamonds denote preferences of respondents who do not hold the respective attribute. The dependent variable is an indicator equal to 1 if respondent i accepts a profile j as a potential partner and 0 otherwise. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

Figure A13: Preference Estimates with Recoded *Maybe*-Responses



Note: This figure shows preferences for network partners when coding *maybe*-responses as indicating acceptance rather than rejection of a profile. Based on equation (3), coefficients represent changes in the probability of a profile being selected as a potential network partner. The red estimates indicate preferences of respondents that hold the respective attribute. Blue diamonds denote preferences of respondents who do not hold the respective attribute. The dependent variable is an indicator equal to 1 if respondent i accepts a profile j as a potential partner and 0 otherwise. The blue bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.

Figure A14: "Sorting" into network connections on preferences



Note: This figure shows estimated preferences for network partners by whether students' stated networks in the survey contain individuals that have or do not have these attributes. The bars reflect 95% confidence intervals. Standard errors are clustered at the individual level.