Dating Experiment

Kalvin, Dan, and Jennifer

December 3, 2017

## Abstract

The idea of self-presentation at the early stages of relationships has been well studied in human interactions and via electronic communications. Our experiment continues this research on the impact of self-presentation of education attainment on online dating success on Tinder. Our research found that ..

## Introduction

Since it’s release in September 2012, Tinder has experienced a meteoric rise from TechCrunch’s Crunchie Award for “Best New Startup of 2013” [[1]](#footnote-23) to being a world-wide dating application with over 50 million users. As of April 2015, Tinder users swipe through 1.6 billion profiles and make more than 26 million matches per day. The application is most popular with young adults with 83% of the user base being under the age of 35. Despite the immense popularity of the platform, there are very few published studies regarding factors that impact Tinder match rate success. Our research begins with a brief overview of the use of technology in dating and builds upon prior research regarding self-presentation in the early stages of relationships. Specifically, we study the impact of education attainment on Tindermatch rate success.

### Brief History of Online Dating

While many believe that online dating is a new phenomenon propelled by the launch of Match.com in 1995, the true conception of the idea goes back hundreds of years. In fact, the first known use of a person using new communication tools to find love and companionship was in 1695 when a 30-year-old British bachelor placed a personal ad seeking “some good young gentlewoman with the fortune of 3,000 pounds or thereabouts.” The movement began with the upper echelons of society and moved to the general populous in the mid 1800s, however it was still viewed as deviant behavior. The expansions of the American frontier further popularized personal ads, however, they became mainstream as lonely World War I soldiers sought female pen pals.

As technology advanced through the mid-twentieth century, dating services began to develop sophisticated methodology for connecting potential suitors. In 1965, a group of Harvard students launched Operation Match, a $3 dating service in which users would submit paper questionnaires that would be processed on an IBM 1401 mainframe computer. The computer matched the users to five potential matches and the results were returned by mail. By the fall of 1965, the business had tens of thousands of users and offices throughout the country.

The 1980s brought about the rise of the VCR and the incredibly awkward era of video dating Suitors would submit a personal video to dating services describing themselves and what they were seeking in a partner. The service would then review the videos and manually match clients by common interest.

The 1990s saw the development of the Internet with the world’s first website and server going live at CERN on December 20, 1990. Within five years, there were 25 million internet users in the United States- that number grew to over 270 million users by 2015. The combination of the new communications technology with the timeless human search for relationships led to the launch of Match.com in 1995. The success of online dating was instantly obvious and in 2002, Wired Magazine prophetically stated, “Twenty years from now, the idea that someone looking for love won’t look for it online will be silly, akin to skipping the card catalog to instead wander the stacks because the right books are found only by accident.” Since that article was written, more than 49 million Americans have tried online dating and approximately 17% of marriages in the last year were products of online dating.

### Prior Studies

Self-presentation and self-disclosure are well studied aspects of the relational development. In Erving Goffman’s 1959 study “Presentation of Self In Everyday Life”, Goffman explores the way way in which individuals use strategic activities to “convey an impression to others which it is in his interest to convey.” These strategic activities are most important at the beginning of a relationship where an individual will “alter their self-presentational behavior in accordance with the values desired by the prospective date.” As Rowatt, Cunningham, & Druen noted, “men, but no women, chang[ed] their self-reported personality characteristics and physical appearance when they expected to meet a potential date. Additionally, their propensity to exaggerate these characteristics was enhanced when the method of meeting was via email.” Dating success, however, is not solely tied to an exaggerated presentation of the ideal self. Reis and Shaver’s research indicated that the need to highlight one’s positive attributes are experienced in tandem with the need to present’ one’s true self. Interestingly, research by Gibbs, Ellison, and Heino indicates that 94% individuals disagreed that they had intentionally misrepresented themselves in their online communication and 87% felt that intentionally misrepresenting one’s self to was unacceptable.

Somewhat contradicting prior research, 2010 research by Hitsch, Hortaçsu, and Airely found no evidence for strategic behavior. Their research found strong same-race preferences among users, which did not differ across age, income, or education levels. The study, however, did show gender differences in mate preferences, specifically that women had stronger preference for income and men had stronger preference for physical attributes.

Our platform of interest, Tinder, has been the focus of a few published studies. Gatter & Hodkinson found that “despite common stereotypes about those who use different types of online dating…no differences were found in motivations, suggesting that people may use both Online Dating Agencies and Tinder for similar reasons.” in specifying the motivations behind Tinder use, Sumter, Vandenbosch, and Ligtenberg found six motivating factors for use: Love, Casual Sex, Ease of Communication, Self-Worth Validation, Thrill of Excitement, and Trendiness.

## Research Design (using ROXO grammar)

Our experiment created two fictitious online profiles that mimicked actual user profiles on Tinder. The profiles were of a male and a female, each 29 years of age. The profiles included five pictures of each individual in various settings, the individual’s name, age, level of education attainment, and a comment stating, “Moving in couple of weeks and looking forward to meeting new people!”

The treatment classes are different levels of educational attainment. Specifically, whether the individual is has earned a PhD, MD, or a Bachelor’s Degree. The control group displayed no educational attainment level in the profile.

The 29 year old individual was selected as it is an age when an individual could plausibly have completed any of the treatment levels of education and it is an age where users were more likely to be looking for long-term relationships, rather than short-term flings.

Based upon prior research, physical attributes tend to be highly significant in partner selection, especially for males, and therefore a single male and a single female were selected for the profile images in order to eliminate variation in match rate based upon aesthetics. Further, the images used were the same images the individuals had used in their actual Tinder profiles to ensure validity of image selection.

The factitious profiles were displayed over a four week period in the eight largest cities in the United States: New York, Los Angeles, Chicago, Houston, Phoenix, Philadelphia, San Antonio, and San Diego (CITE ME). It is important to note that neither individual whose pictures we used had previously used Tinder in any of the test cities, thus reducing the potential for any contamination or bias in the results. Each profile-educational level combination was displayed to 100 users in each city for a one week period of time in November-December 2017. Test subjects were randomly selected by “swiping right” on every other profile displayed. It should be noted that Tinder profile viewership was restricted to only those who were chosen as test subjects. This eliminated the ability for a test subject to view the profile at multiple levels of education. The only potential for spillover would be by an individual having multiple profiles and having been selected twice as a test subject or by recognizing the profile while viewing Tinder on another user’s account.

After one week, we returned to the profile to collect data on which users also “swiped right” on our test profile. This factorial design of two sexes, eight cities, and four education levels allowed for the analysis of different impacts of test conditions across sexes and locations.

Linearly, the design was: R - - 01, R X1 - 02, R X2 - 03, R X3 - 04

## Randomization engineering

Gathering a larger sample of profiles gave our analysis more power and confidence in the trends observed from the data <>. The pilot demonstrated that mannually copying and pasting the details from each suitor’s profile prohibited collecting a larger number of profiles. It took about three hours to gather four hundred profiles, meaning just over 2 minutes per profile. The use of automation was a very effective tool that enabled the collection of over five thousand suitor profiles in a four week period, which would have taken one hundred eight five hours by hand. Luckily, the website version of Tinder exactly matched the mobile phone version, greatly streamlining and standardizing the data collection approach.

The research team wrote scripts in AutoHotKey (PC) and Maestro (the MAC equivalent) that control the operating system of the computer. The code would open up Tinder and navigate to developers’ tools where the website’s Javascript rendered soruce code (the HTML behind the webpage) was available. The source code listed the suitors name, age, a unique photo source path (very helpful for data validation later), and two details about the profile. After collecting and storing the source code in the appropriate folder that designated the city and treatment level, the code would automatically right swipe on the profile collect.

Tinder restricts users from seeing repeat profiles to prevent communication between people who previously interacted unsuccessfully, making the trials independent. Each week a new randomly select batch of ladies and gentlemen were swiped on and put into the experiment. Any user of the Tinder application is subject to the behind the scences algorithm that serves up potential profiles; both the test male and female profiles, along with the profiles of the general Tinder user.

To further enhance the level of randomness within treatment assignment, every other profile was systematically rejected from the experiment. Given more time, the treatments could be delivered in alternative orders to help determine if the algorithm modifies the suitor population based on features of the profile. As the experiment was constrained to a five week window, the collection of covariate distributions augments the research team’s understanding of the mechanisms driving tinder matches. For example, to help demonstrate the balance in treatment assignment, the distribution of suitor ages for each treatment condition are displayed in figures XX-XX below.

Figure XX (will add plots after they’re finalized– please check EDA google doc)

It is interesting to see that the age distribution has a right skew for all treatment conditions, and it is unclear that depending on Tinder to select suitors is truly random– it could be the case that Tinder selects suitors based on the age set in the test profile, and it could also be the case that the general population of Tinder users is centered around the mid-20s age range. But the consistency of this distribution across all conditions indicates that there was indeed an equal probability of receiving a given treatment condition, at least amongst the suitors selected by Tinder.

The age distribution of suitors to the Male profile (female suitors) does seem to be centered approximately 2 years younger than that of suitors to the Female profile, and the skew appears to increase slightly with increasing education level. However, statistical tests for the difference in average age by test condition yields a significant difference only between the male and female conditions (shown in figures XX-XX below), and its effect size is small– approximately half a year.

## Experimental materials (e.g. treatment materials)

Objectively measuring individual’s success in life is ambiguous and difficult, but a reasonably good proxy for research purposes could be educational attainment, job title prestige or educational institution prestige. Below is a table of potential example degrees of severity for each of the proposed measures of success:

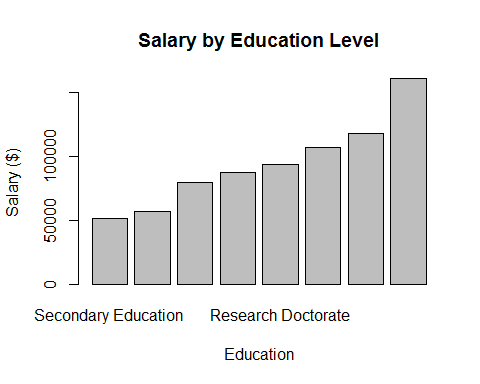
|  |  |  |
| --- | --- | --- |
| Educational Attainment | Job Title | | Educational Prestige (nearest) |
| Advanced Degree (JD, PhD, MD, etc) | Doctor | | IVY (Harvard) |
| College Degree | Teacher | State School (UMass) |
| Associate Degree | Social Worker | Community College (Bunker Hill CC) |

As a team we considered the strengths and weakness of each respective success measure. We planned on customizing the education institution to reflect broadly known schools in the vicinity of each city in the experiment. To enhance consistency of school selection across geography, Barron’s Ranking list and the US News school ranking would have been consulted. However one of the drawbacks of educational prestige, that could not be remedies, is that educational prestige is most often determined while the individual is in their early twenties. At such a young age the future of any person is extremely malleable and fertility concerns are not particularly potent. Generally it also takes time to become particularly skilled or develop expertise or to develop a positive reputation in a domain of interest. The research team leaned away from educational prestige as a success measure because it would not get close to measuring a potential tradeoff between prioritizing relationships and family in lieu of pursing other interests tied to XXX (quantifiable success ??) <>.

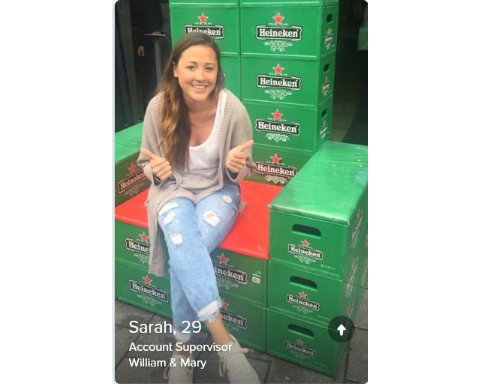
Professions then appeared as the better measure for success as individual’s career paths are more fleshed out a few years later. Most adults know that doctors make significantly more than teachers who make more than social workers, however all three professions are in the business of helping people, and this could introduce potential bias. To elaborate on that point about bias, men might seek women with a nurturing profession (picturing the woman raising babies at home) and women might avoid men for fear of being replaced in the traditional household roles. Additionally Researchers at Microsoft recently published a paper demonstrating gender based job title bias in the general public’s vernacular, implying that this success measure potentially introduces bias into an experiment (Bolukbasi, 2016). Ultimately professions each carry their own reputation and predefined characteristics and controlling for that would be exceptionally difficult. Additionally, the novelty effect for women and men in unconventional roles also raised some concerns. Cross comparing average salary data from the Occupation Employment Statistics survey and the Current Population Survey reveal that top paying professions are male dominated. For example, the Bureau of Labor Statistics data shows that 66% of dentists, 64% of lawyers, and 73% of CEOs are men. Meaning that any high paying job selected would be more common among male profiles than female profiles making an apples to apples cross gender comparison questionable. For these reasons the final success measurement was not professions.

The government and society have created very clear breaks in educational attainment that are known to all, uniform across geographic lines, and would be easy to administer for an experiment. Generally people perceive access to education in the United States as exceptionally balanced across gender and bias concerns are mitigated. As the graph below indicates, on average there is a positive association between higher education and income levels, providing some validation for the positive life impact. After deciding to run the experiment with educational attainment as the measure of success, the degrees of treatment were identified as MD, BS, and PhD. The research team was particularly interested in the difference between the impact of having a Medical Degree versus having a Research Degree. Both degrees require about five years to obtain, however there is significant difference in earnings potential. Due to concerns regarding the realness of the profile and limited time, the team decided not to test the impact of the associate’s degree because other profiles on Tinder did not prominently display and associates degree. The control group would be a profile giving no information for the particular success measure. The treatments were delivered in the following non-monotonic order: MD, BS, no education, and PhD.

In order to effectively deliver the treatment variable, showing the profile to a limited audience of potential suitors was paramount. To elaborate on that point, should a potential suitor see the profile change from “MD” to “BS” that would not only ruin the data point, but also risk the injured party flagging the account to the administrators and terminating the experiment prematurely. Fortunately, the premium Tinder account enabled the research team to control who had access to the profile so only the one hundred individuals in each city selected for that week’s treatment were exposed; very effectively reducing potential spillover.



The potential suitors were shown a profile that either read: “MD”, “PhD”, “BS” or the field to provide education information was left blank. A sample profile can be seen below.



## Measurement of variables

As discussed previously, the treatment in this experiment is the test profile’s exposure of a given education level to a potential suitor, and the outcome is whether or not the suitor matches with the test profile. When applying an education level treatment to the test profile, the measured test subject covariates include its age, sex, and location, and the measured suitor covariates include age, sex, whether or not school information is provided, whether or not job information is provided, and whether or not instagram photos are provided (and their count, if so). These variables are described in table below.

## Currently generic markdown table using pandoc is not supported.

Table XX: Description of Variables

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Variable Type | Source | Description / Possible Values |
| test profile education level | treatment | test profile | 4 possible values\*: No education listed (control), Bachelor’s, MD, PhD |
| suitor profile match | outcome | Tinder application | 2 possible values: match or no match |
| test profile sex | covariate | test profile | 2 possible values: female or male |
| test profile location | covariate | test profile | 8 possible values: Chicago, Houston, Los Angeles, New York, Philadelphia, Phoenix, San Antonio, San Diego |
| suitor age | covariate | suitor source code | Due to the filters set in the test profiles, this variable can range from 24 - 34 years. |
| suitor school\*\*\* | covariate | suitor source code | 2 possible values: school information was detected or not |
| suitor job\*\*\* | covariate | suitor source code | 2 possible values: job information was detected or not |
| suitor instagram\*\*\* | covariate | suitor source code | 2 possible values: suitor profile included an instagram link or not |
| number of instagram photos\*\*\* | covariate | suitor source code | if a suitor profile contained a link to an instagram account, this variable is the number of photos in the account |
| \* The initial experimental plan | also included t | he ‘Associates’ educa | tion level as a treatment condition, however, technical issues at the start of the experiment prevented proper execution of the ‘PhD’ education level treatment, so the ‘PhD’ education level treatment was repeated and the ‘Associates’ education level treatment was skipped. |

\*\* The test profile age was an originally planned covariate, but due to time constraints, only one age was used for the entire experiment.

\*\*\* These variables are believed to be only partially observed– see Section for further details.

*Issues with Measurement of Outcome*

After swiping right on a suitor, the Tinder application only provides notification if that suitor likes the test profile in return, and does not provide an indication of whether or not a suitor actually saw the profile or if the suitor disliked the test profile. Our experiment therefore contains some rate of non-compliance and we are unfortunately unable to determine the compliance rate– our estimated average treatment effect is thus the effect of the intent to treat.

*Issues with Measurement of Covariates*

The structure of the relevant source code in the Tinder browser interface is such that it contains a link to a suitor’s profile image (which we use as a unique identifier for that suitor), the suitor’s name, the suitor’s age, and finally a maximum of two additional details that may include any combination of school information, job information, and instagram information.

It is known that a Tinder profile may display all three of these pieces of information, but none of the ## source codes collected contained more than two details. It is highly unlikely that amongst the ## suitors encountered in the experiment, none displayed more than two of these details, so we believe that school information, job information, and instagram information are only partially observed, and without more in-depth examination of each suitor profile at the time of experiment execution, our analysis is unable to fully determine the education and job status/level, as well as instagram information, for each suitor. Due to time constraints, thorough examination of each profile was unfortunately not an option– we were required to use an automated swiping method which acquired covariate information from source code. It is also known that a Tinder profile may additionally display the suitor’s distance and favorite spotify songs, but none of this information was present in the suitor source codes acquired during the experiment, and these two covariates were thus unobserved.

The source code indicates which suitor details are the profile image, the name, and the age, but does not indicate which details are school, job, or instagram information. Our data collection process searches the text of these details to make a best guess as to what type of detail it is (instagram information is easily identified, but not school or job information). So for some of the school and job details that were collected, it could not be determined whether the detail represented education or employment.

*Data Pre-processing* Since the non-binary variables are categorical, they were binarized using indicator variables for each category…

## Modeling choices

calculation of statistical power

## Experiment Results

In text description of your results Figures and tables that support your in text description Clean, clear, well articulated relationships between your theory, your hypotheses, the numbers that your models produce, and the figures you present

1. <https://techcrunch.com/video/tinder-wins-best-new-startup-of-2013-crunchies-awards-2013/518118930/> [↑](#footnote-ref-23)