Effect of Profile Pictures on LinkedIn Connection Acceptance Rate

W241 Final Report | Fall 2023

Mahmoud Ghanem, Mark Haase, Hope Huang, Andrew Sandico, Abraham Yang

December 21, 2023

Introduction

In the age of professional networking, online platforms like LinkedIn play a crucial role in connecting individuals across diverse industries. Studies have shown that it takes only one-tenth of a second for someone to form impressions about you based on your photo (Wargo). Research indicates that 93% of recruiters will examine your profile photo before reaching out to you (Hunt). As professionals increasingly rely on virtual interactions to expand their networks, understanding the dynamics of visual representation in a digital professional space becomes vital. Does having a LinkedIn profile picture actually cause an increase in invitation acceptance rate? This study investigates the impact LinkedIn profile pictures may have on the acceptance rates of connection invitations.

To conduct the experiment, ten identical profiles were created with each having either a profile picture or not having a profile picture. Four accounts were female profiles and six accounts were male profiles. With the data left available, it was determined that the accounts with profile pictures had a connection acceptance rate of 14.6 percentage points higher than the accounts without profile pictures. This study could benefit from further research and exploration of different demographics and specific characteristics within profile pictures such as attire and facial expressions.

Research Question

We are interested in quantifying how large of an effect profile pictures play in accepting a connection request from a stranger on LinkedIn. Do LinkedIn profiles with profile pictures yield higher connection acceptance rates?

Hypothesis

We expect the treatment group who set up LinkedIn profile pictures when attempting to connect with new people receives a higher connection acceptance rate. Firstly, this may be because of the power of photographs. A new study published in *Social Psychological and Personality Science* found that someone's first impressions of you from a photograph are likely to stick, even after you meet in-person (Gunaydin). Secondly, LinkedIn profiles with photos receive more views. LinkedIn, the company itself, states on profiles with no picture that, “members with profile photos can receive up to 21 times more profile views than those without profile photos” (Quantity Improvement Solutions). Therefore, it is fair to assume that the higher the amount of profile views, the more likely someone connects with you.

The experiment aims to assess the above hypothesis that LinkedIn profiles with profile pictures receive statistically significantly higher connection acceptance rates than the LinkedIn profiles without profile pictures.

Experimental Design

Treatment Design

Our experiment measures the effect of having a picture on your LinkedIn profile on the likelihood that others will accept your invitations to connect with them. As a secondary factor, we also determine the effect of the sender being male or female on the likelihood that others accept the invitation. The subjects (or recipients) of our experiment are randomly assigned to receive the picture treatment (pic / no pic). If the user sees the invite, they can click an “Accept” button to accept the invite or “Ignore” button to reject the invite. LinkedIn invitations allows a sender to include a note in their invitation, but to keep the experimental design clean, we did not include notes with our invitations. This project is a randomized experiment using a posttest control group design which consists of a randomized allocation to groups (R), Treatment (X), no treatment (-), and observation or measurement. A posttest design means that the analysis is completed after the treatment. In other words, there was no pretest. The group with treatment is the experimental group and the no treatment group is represented below in the control group.

Experimental group	R X O
Control group	R - O

Figure 1 - ROXO Diagram

Randomization

Subjects for this study were part of a LinkedIn group called “Analytics and Data Science Career” which consisted of 289,000+ members during the week of November 3rd when we drew our sample. This group was specifically selected due to the 10 profiles having job titles and education based on the professionals in the group. This approach would increase the relevancy for subjects/recipients to accept any invite from the profiles.

Due to LinkedIn restrictions of not being able to access the full list of 289,000+ members, the most recent 2,500 members of the group were scraped from the site. This number of 2,500 was selected to help meet the power analysis threshold of 80% (see power analysis section). After the 2,500 was scraped, a random integer was assigned to each subject/recipient. The random integer was then sorted in ascending order, and we assigned blocks of 500 recipients to each team member. The process was repeated for an additional 328 members as they appeared in the new members queue which resulted in 2,828 subjects/recipients. This approach ensured only one invite request from each profile to a subject and created clarity for team members to send out an invite to.

Measurement Design

A direct single measurement design was used for this project. We created pairs of treatment and control profiles on LinkedIn. Each treatment profile has a picture, while each control profile does not have a picture (see Figure 2). Each pair of profiles is identical in terms of education background and job history. The profiles in each pair different only in terms of having picture/no picture and in their first names. We chose different first names in order to reduce the likelihood of being identified as inauthentic accounts, however we did choose stereotypically caucasian names to avoid introducing differences that might bias our results, such as perceived nationality or ethnicity. We created 10 profiles, with 6 being male and 4 being female.

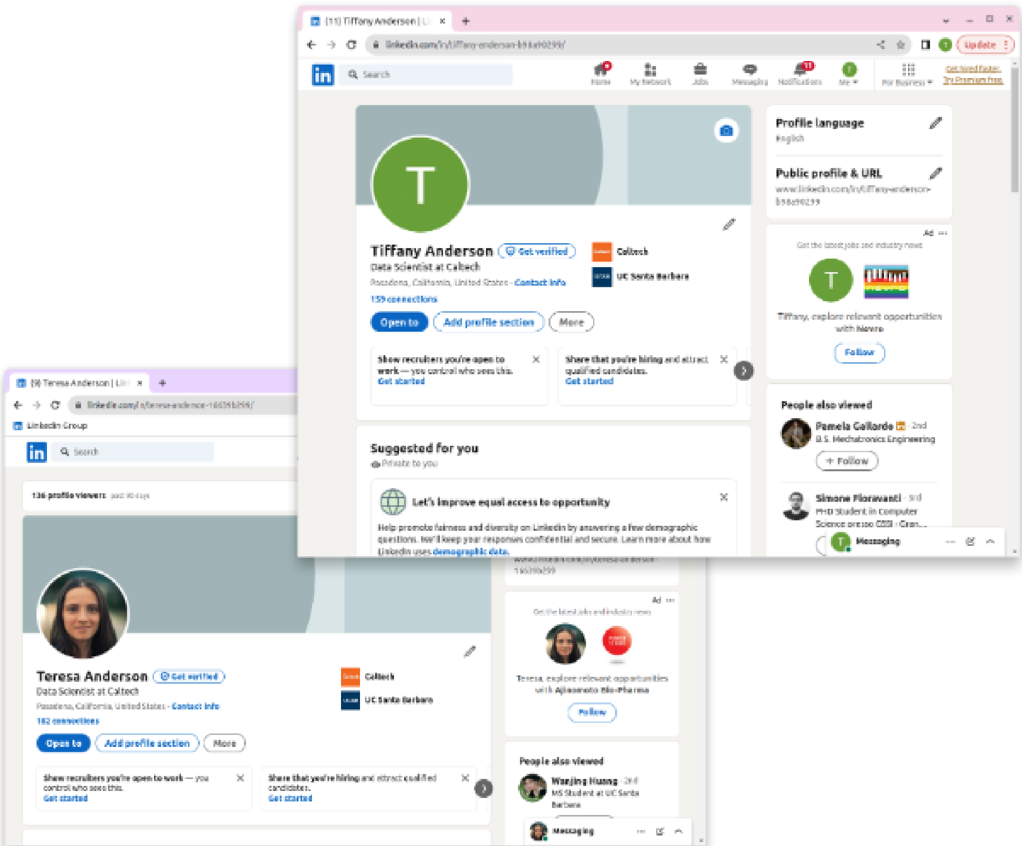

Figure 2 - A pair of profiles that vary only in the treatment condition (picture or no picture) and first name

Invitations were sent out over a few weeks during the dates of 11/3 to 12/7. When an invitation was accepted on LinkedIn, each profile was able to confirm the connection under the my network section (see Figure 3).

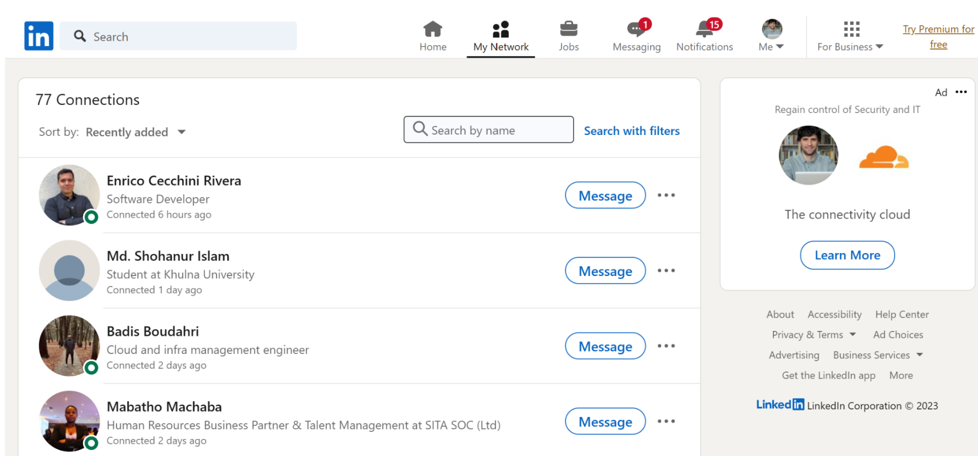

Figure 3 - Viewing which recipients accepted the invitation

In order to avoid biasing our estimates due to temporal confounders (e.g. certain days of the week are more likely to yield accepted invitations), we spaced out the sending of invitations into small batches sent over the course of several weeks, including different days of the week and different times of day. Another factor that contributed to this batch process of invites was that LinkedIn places limits on how many invites a profile can send. In particular, there are limits on how many invitations can be sent per week (30 maximum) as well as how many open invitations you can have at one time. Due to the weekly limit, we distributed our invitations across 10 LinkedIn profiles and sent a small number (10-15) of invites per day.

On LinkedIn, the sender of an invite cannot distinguish between invitations where the recipient clicked the "Ignore" button, recipients who saw the invite and did not click any button, and recipients who did not even see the invite. Therefore we have a source of non-compliance (somebody who rarely checks their LinkedIn and does not see our invite) as well as two similar outcomes (clicking the "Ignore" button versus literally ignoring the invitation). For the purposes of this experiment, we consider a "intent to treat" causal effect since we do not have the ability to measure non-compliance. If the recipient accepts the invitation during the duration of our multi-week experiment, we code that as a 1. All other outcomes are coded as 0.

To limit confounding factors that may occur with our process, the following process was used to ensure consistency across 10 profiles and 5 team members:

- Each team member was assigned a pair of treatment/control profiles. That team member only managed those two profiles.
- We selected profile photos from actors' headshots on Shutterfly and cropped them as necessary to look like reasonable profile photos.
- Each team member committed to send requests in small batches from November 3 through December 7.
- Our goal was to send the maximum amount of invites per profile. Due to LinkedIn invitation limits, this turned out to be about 80-90 invitations per profile per week, although in some cases, our accounts were flagged and required us to send fewer invitations than we had hoped for.
- Invites were only sent to subjects that were part of the Analytics and Data Science Career group in LinkedIn [Analytics and Data Science Career] at the start of the randomization process.
- We ensured that each recipient receives only one invitation request from our study.

Experiment Flow Diagram

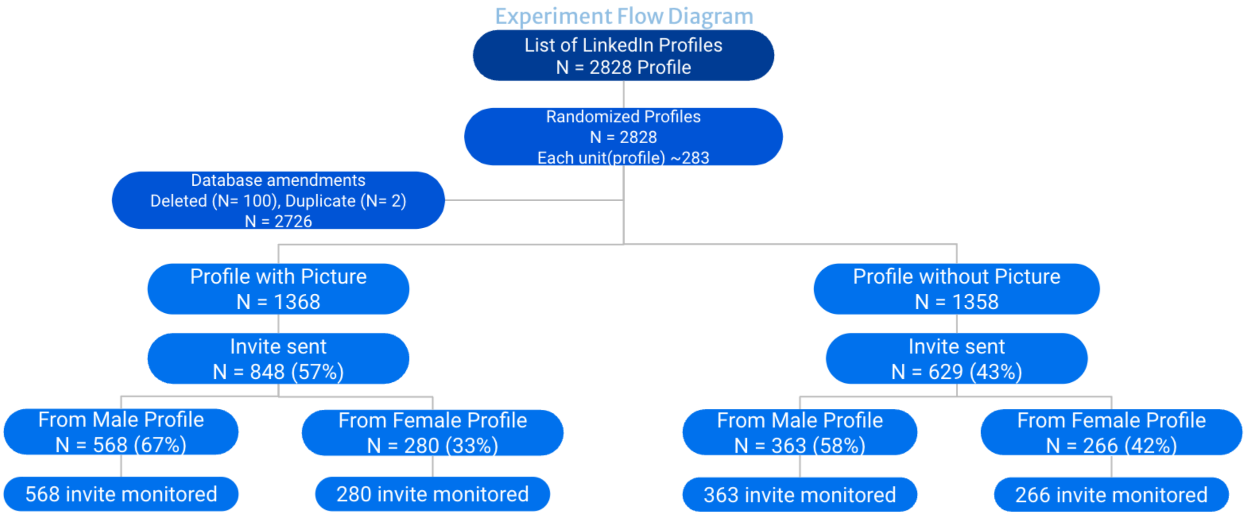

Figure 4 - Experiment Flow Diagram

The flow diagram (Figure 4) shows how we scraped data from **2,828** LinkedIn users belonging to a data science group. Each user was randomly assigned to either the treatment group (5 profiles with pictures) or the control group (5 profiles without pictures). The **2,828** LinkedIn users were randomly assigned to each one of the 10 profiles, approximately 283 users to each profile.

After data cleaning, which involved removing duplicate profiles and attrited users, the final population comprised **2,726** users and were split between the treatment group (**1,368** users with pictures) and the control group (**1,358** users without pictures).
Out of the **1,477** users who received invitations:

- The treatment group **848** users (57%) received invites consisted of:
    - **568** (67%) Males
    - **280** (33%) Females
- The control group **629** users (43%) received invites consisted of:
    - **363** (58%) Males
    - **266** (42%) Females

Power Analysis

We conducted a pre-experiment power analysis to assess the ability of our experimental design to detect significant effects in various scenarios related to LinkedIn connection acceptance. We simulated three scenarios, each representing different conditions. We assumed a 5% treatment effect inspired by studies in the field.

In the first scenario, we defined the treatment based on whether a LinkedIn profile included a picture. Through the generation of random data and repeated simulations with incrementally growing sample sizes, we determined that an approximate minimum of 2,400 subjects is required to achieve 80% power to detect a treatment effect at a significance level of 0.05.

The second scenario maintained the same treatment effect as the first scenario but introduced larger variance in the treatment effect. The results indicated a similar sample size requirement of around 2,800 subjects for 80% power, demonstrating the robustness of the analysis.

The third scenario expanded the experiment to a multifactorial design, incorporating both the presence of a profile picture and the gender of the sender profile. This more complex scenario, accounting for main and interaction effects, suggested a larger sample size requirement of approximately 7,000-8,000 subjects to achieve 80% power for detecting any significant effects.

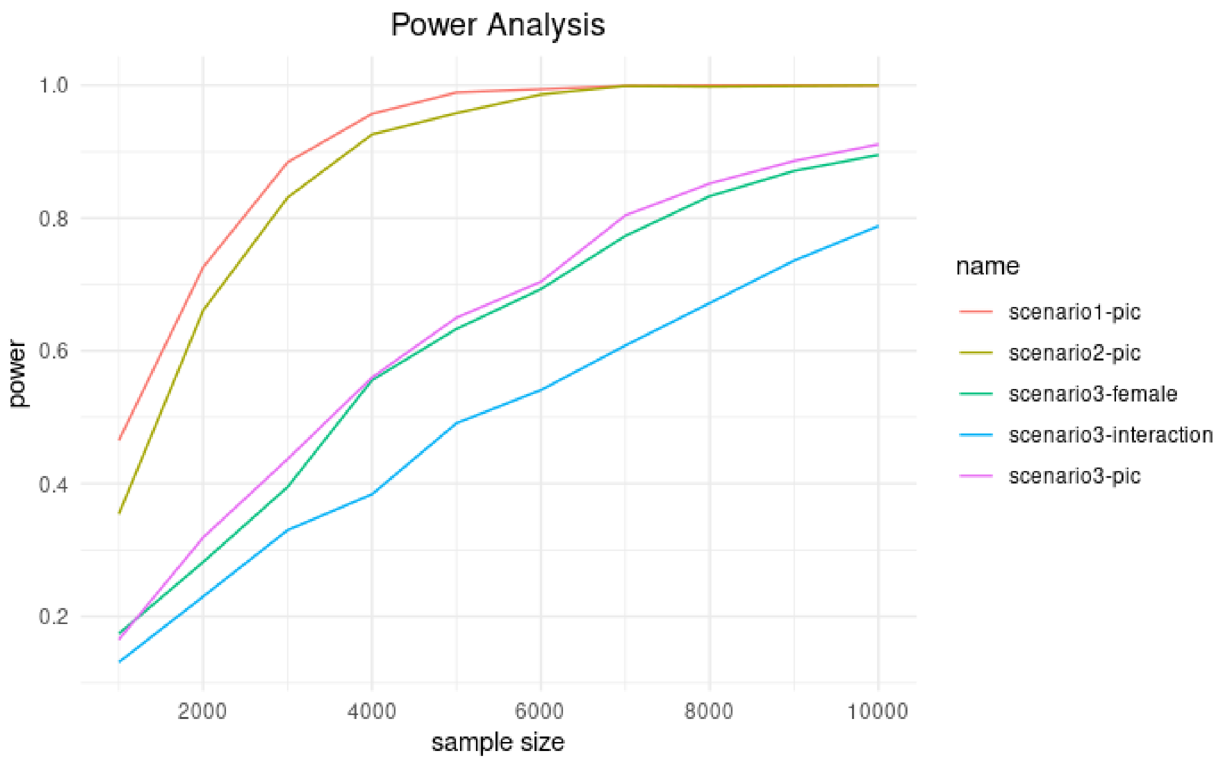

Figure 5 - Different Scenario Power Analysis Chart

In summary, our power analysis provided valuable insights into the sample size needed to achieve a desired level of statistical power for each experimental scenario, aiding in the planning and design of the experiment and guided us to target approximately 2500 sample size to achieve the desired level of statistical power for detecting significant effects. The experiment observed higher than anticipated acceptance rates in the multifactorial experiment, especially for profiles with pictures, contribute to the current sample size of 1477 being sufficient for detecting significant effects. The unexpected effect sizes, particularly in male profiles with pictures, imply a more pronounced impact than initially estimated. As the actual effect sizes are larger than assumed, the experiment gains increased statistical power, allowing for the detection of significant effects with a smaller sample size.

Factorial Design

Treatment	Female	Male	
Picture	175	372	273.5
No Picture	155	159	157
	165	265.5	

Figure 6 - 2 x 2 Factorial Design

The 2 x 2 factorial design table demonstrates the influence of Gender and Picture on invitation acceptance. It reveals that the impact of having a profile picture on invitation acceptance varies based on the user's gender, suggesting an interaction effect.

Figure 7 - Interaction effect chart for treatment

The Interaction effect graph highlights the performance discrepancy, emphasizing specific combinations, such as male users with profile pictures, outperform others in invitation acceptance and show that male profiles without pictures have less chance of getting invites accepted. Male and female profiles without picture acceptance rate are close to each other.

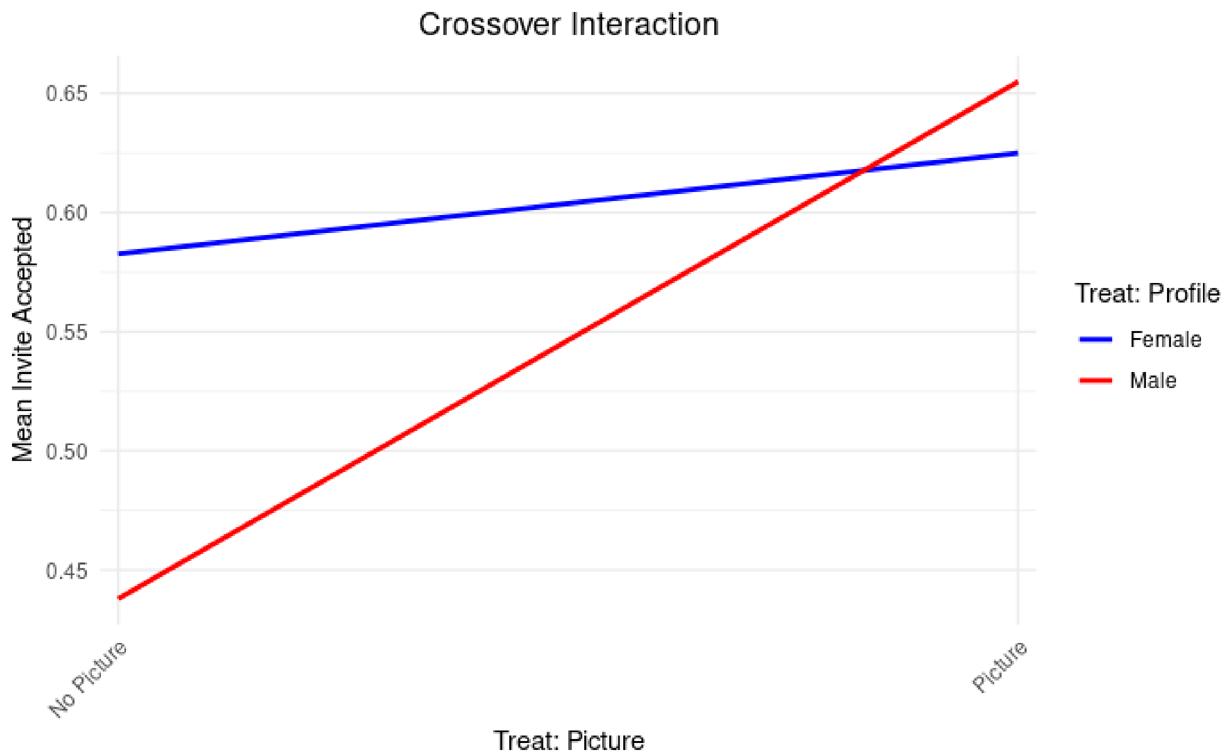

Figure 8 - Crossover Interaction Effects between treatment

The cross-over interaction graphs emphasize the nuanced nature of the interaction effect, showing that acceptance rate varies based on the interaction of both gender and profile picture.

Difficulties Collecting Data

For the most part, data collection went well. Four of our five team members were able to send out LinkedIn requests without encountering any issues. One team member had all of his LinkedIn accounts for the experiment blocked shortly after creation. We used various tactics to attempt to bypass these countermeasures, although these were not 100% effective in all cases. The tactics used involved creating new Gmail accounts, registering real phone numbers from friends and family, using the Berkeley VPN to disguise the source of the traffic, using dedicated web browsers in incognito mode, clearing all cache and cookies, limiting the number of invitations sent at the onset of the account, and clicking on advertisements.

Another difficulty we encountered was the weekly connection requests limit imposed by LinkedIn. LinkedIn limits accounts to sending no more than a certain number of requests per week. During the study, certain accounts were limited to 30 weekly invites while others were 100. A majority of the non-profile pictures were limited to 30 weekly invites. These limit caps were unexpected and caused our approach to change from the initial planning and power analysis, which is a change of a goal of 2,500 observations to 1,477.

Furthermore, we encountered attrition where recipient accounts disappeared. This could have been due to name changes, members leaving the professional group, or terminating their LinkedIn account altogether. We added an attribute to our dataset to encode for attrition.

Experimental Results

After accounting for our banned profiles as well as subjects' attrition, we ended up sending 1,477 LinkedIn invitations during the course of the experiment. The data contains two binary treatment variables: “Treat Picture” means that an invitation was sent from a LinkedIn profile that has a profile picture, and “Treat Male” means that an invitation was sent from a LinkedIn profile with a male persona, i.e. a male name (and a male picture, for the profiles that have pictures). We have one binary outcome variable, which is whether the recipient accepted the invitation.

To increase precision, we also collected several covariates for each subject. The first covariate is a binary indicator for whether the account has been inactive for 6 or more months (Figure 9). (Activity is defined as having posted, commented, or reacted to a post.) We included this covariate in the expectation that some accounts might be infrequently used, dormant, etc.

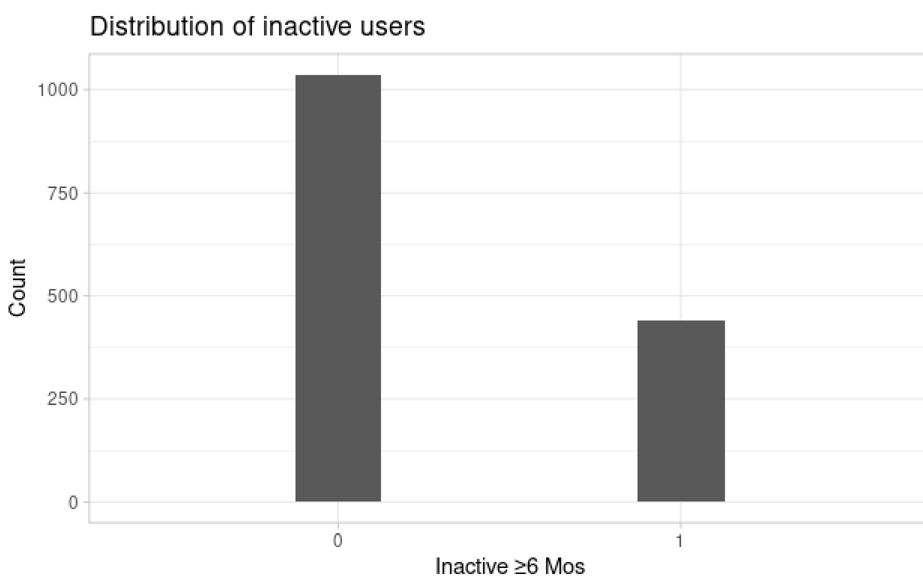

Figure 9 - Distribution of Inactive Status for Experimental Subjects

The second covariate is “Invite Duration (Days)”, which represents how many days passed between when the invite was sent and when the sending account was last checked. We originally included this value because we were concerned if we sent invitations shortly before the experiment ended, we might not have enough time to see the response, even if the recipient ultimately did accept it. On the other hand, invitations that were sent a week or two before the end of the experiment would have much more time for the recipient to see it and act on it. This covariate was even more useful than we initially anticipated, because several of our accounts were banned midway through the experiment. By including this covariate, we are able to control for the fact that some invitations were sent just shortly before the ban was enacted.

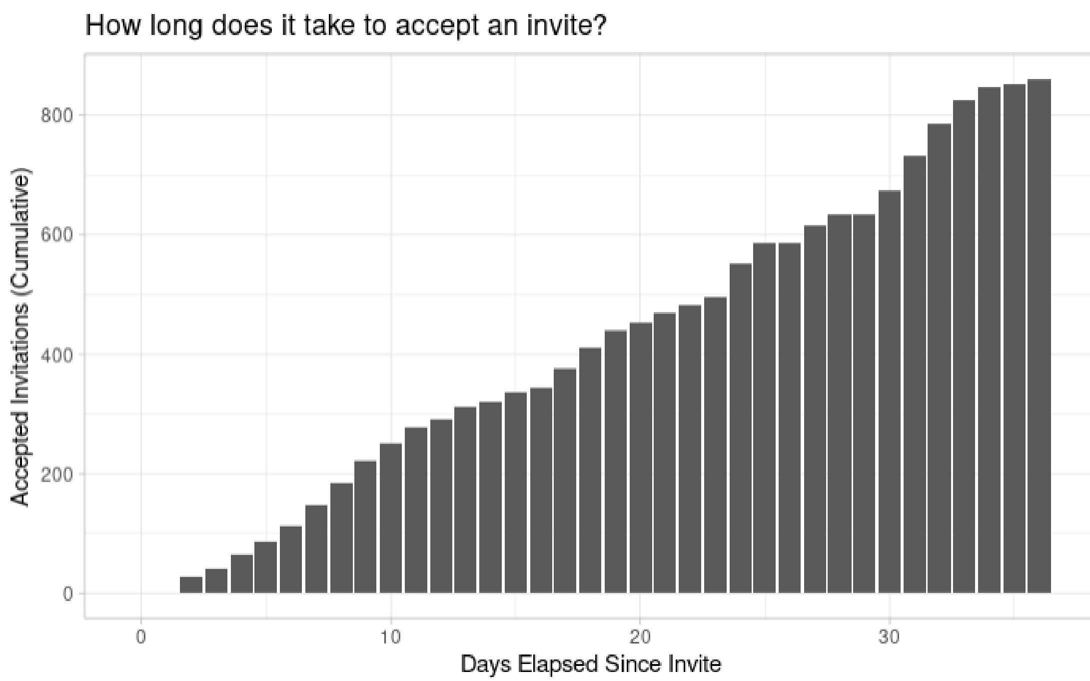

Figure 10 - The cumulative number of invitations accepted grows approximately linearly in the number of days the recipient has to respond to the invitation

We plotted the cumulative count of accepted invitations against the invite duration (Figure 10). We were concerned that the distribution might have some non-linear shape, e.g. logarithmic decay. The plot looks linear, however, and so we included this as a linear term in our models.

The third covariate is the perceived gender of the *recipient* of the invitation (Figure 11). This data is a bit tricky to collect. Our team recorded the perceived gender of the recipient based on name, profile picture (if they had one), or–in rare cases–gendered pronouns listed on their profile. When gender could not be determined, we coded the gender as “na”. This coding might introduce some bias, for example, we were more likely to code “na” gender for foreign names where we were not familiar with the gender implied by a name. Also there may be some correlation between the recipient's culture and their choice not to post a profile photo (e.g. a cultural norm in a specific country that women don’t post profile photos) which would also bias this measure.

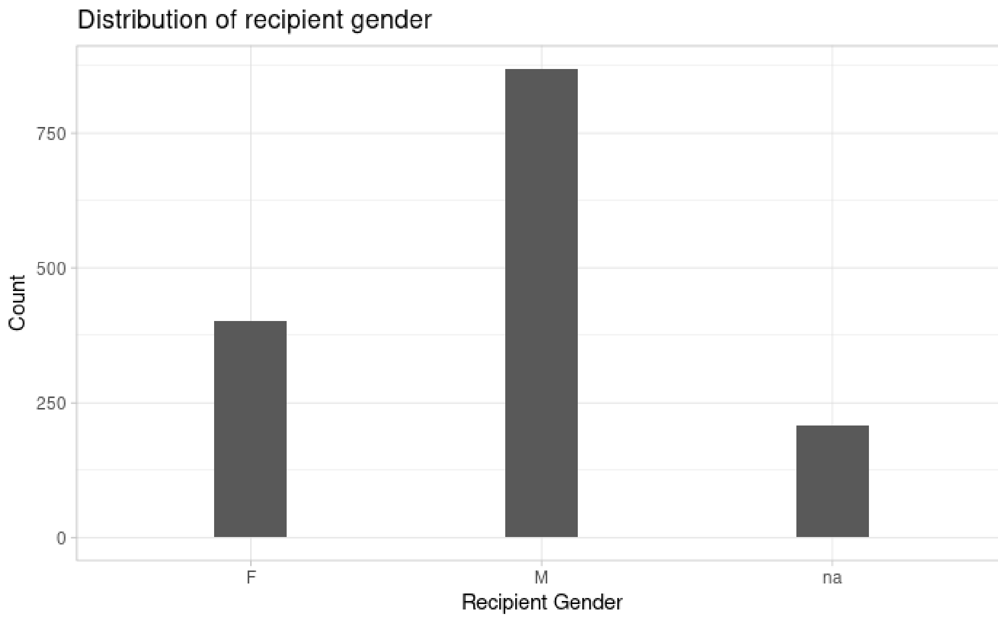

Figure 11 - Distribution of Recipients by Perceived Gender

The team planned out the data schema and most of the data coding prior to beginning the collection of data, which greatly reduced the amount of data cleaning that we needed to do afterward. We recorded our data in a Google spreadsheet, which we exported to CSV and then imported into R as a data table. The minimal data cleaning required included removing attrited subjects, parsing date strings, and converting our start and end dates into elapsed durations.

Regression Models

We constructed four regression models from our experimental data. A summary of the models is displayed in Table 1.

Table 1: LinkedIn Experiment Results

	Dependent variable:			
	Invite Accepted			
	(1)	(2)	(3)	(4)
Treat Picture	0.146***	0.042	0.138***	0.166***
	(0.026)	(0.042)	(0.025)	(0.049)
Treat Male		−0.145***		
		(0.040)		
Treat Picture:Treat Male		0.175***		
		(0.053)		
Inactive 6 Mos			−0.107***	−0.076**
			(0.027)	(0.030)
Invite Duration (Days)			0.008***	0.009***
			(0.001)	(0.001)
Recipient Male				0.027
				(0.045)
Treat Picture:Recipient Male				0.018
				(0.059)
Constant	0.499***	0.583***	0.387***	0.337***
	(0.020)	(0.030)	(0.029)	(0.045)
Observations	1,477	1,477	1,477	1,270
R^2	0.021	0.031	0.061	0.072
Adjusted R^2	0.021	0.029	0.059	0.068
Residual Std. Error	0.488 (df = 1475)	0.486 (df = 1473)	0.478 (df = 1473)	0.475 (df = 1264)
Note:				*p<0.1; **p<0.05; ***p<0.01
				All standard errors are robust standard errors.

The first model includes just the primary treatment effect and an intercept term. We saw a 14.6 point increase in invitation acceptance rate, a surprisingly large effect size that exceeded our pre-experiment assumption of a 5 point effect and is statistically significant.

The second model uses a multifactor design to test the hypothesis that having a profile picture produces a different effect size for male senders than for female senders. The surprising result shows a 14.5 point decrease for male senders without a profile picture but a 3 point *increase* for male senders with a picture. Both effect sizes are statistically significant. However, we also see the statistical significance of the picture treatment go away in this multifactor model, a complication that we address in the next section.

The third model builds on the first model by adding the covariates for “Inactive 6 Mos” and “Invite Duration (Days)”. The coefficient for the inactivity indicator variable reflects a 10.7 point decrease in acceptance rate among inactive recipients. This effect is statistically significant and the negative sign makes sense: less active users are less likely to accept an invitation. The coefficient for invite duration is also statistically significant and the positive makes sense: each additional day that the recipient has to respond to the invitation is associated with a 0.8 point increase in acceptance rate.

Adding these two covariates to the third model does not change the precision of our estimated ATE very much, but the ATE is positive and statistically significant, showing that the picture treatment causes a 13.8 point increase in acceptance rate.

The fourth model includes our covariates plus a heterogeneous treatment effect (HTE) term to investigate if male recipients respond to the picture treatment differently than non-male recipients. This model shows a strong and significant effect for the picture treatment as well as the inactivity and invite duration covariates. It shows a very small HTE for male recipients who receive the picture treatment, but this effect is quite small relative to the treatment effects and it is not statistically significant.

Covariate Balance Tests

Given the concern about randomization noted in the multifactor model above, and seeing that randomization is imbalanced, we ran a covariate balance check. We constructed a null model consisting of regressing the picture treatment against an intercept term, and another model regressing the picture treatment against the covariates “Recipient Male”, “Recipient NA”, and “Inactive 6 Mos”. Then we used an F-test to compare the null and full models.

Table 2: Covariate Balance Test

	Dependent variable:	
	Treat Picture	
	(1)	(2)
Recipient Male		0.021
		(0.030)
Recipient NA		0.002
		(0.043)
Inactive 6 Mos		-0.061^{**}
		(0.028)
Constant	0.574^{***}	0.580^{***}
	(0.013)	(0.027)
Observations	1,478	1,478
R^2	0.000	0.004
Adjusted R^2	0.000	0.002
Residual Std. Error	0.495 (df = 1477)	0.494 (df = 1474)
Note:		*p<0.1; **p<0.05; ***p<0.01
		All standard errors are robust standard errors.

The resulting p-value of 0.135 means we fail to reject the null hypothesis, and therefore no substantial evidence of covariate imbalance.

Conclusion

The accounts with profile pictures were found to have a statistically significant effect of increasing connection acceptance rate by a magnitude of 14.6 percentage point by the simple linear regression model. Therefore, we were able to reject the null hypothesis that the presence of LinkedIn profile pictures does not affect connection acceptance rate. We fail to reject the alternative hypothesis that the accounts with profile pictures receive a higher connection acceptance rate. No heterogeneous treatment effects on connection acceptance rate were found with the gender of the recipient.

From the multi-factor model, we found that the male accounts with profile pictures have a statistically significant effect of increasing connection acceptance rate by a magnitude or 17.5 percentage point while interestingly, male profiles with no profile pictures decrease connection acceptance rate by 14.5 percentage point.

Additionally, two covariates, invite durations and participant activity status, were found to be statistically significant. Each additional day the recipient has to respond increases connection acceptance rate by 0.8 points, while inactive profiles are associated with a 10.7 point decrease in connection acceptance rate. These findings align with general intuition.

Overall, the results of our experiment provide valuable insights to the online professional networking field. This study highlights the significance of profile pictures in the context of building successful connections on LinkedIn.

Limitations and Future Work

While this experiment provides valuable insights into the dynamics of LinkedIn connection acceptance rates, several limitations must be acknowledged. Firstly, the sample used in this study consisted of profiles with names commonly perceived as having a white association (e.g., Ethan Miller). Secondly, the participants of the study were selected from the *Analytics and Data Science Career* LinkedIn group. This targeted recruitment strategy may result in a sample that is not representative of the broader LinkedIn user population. Consequently, the findings from this study may be skewed towards the perspectives prevalent within the data science community and may not capture the connection acceptance rates that profiles of various ethnicities would have received.

In the future, it would be interesting to expand this study with a more diverse demographic of profiles. Investigating the potential variations in the impact of profile pictures on connection acceptance rates across different ethnicities, cultures, names, and industries could reveal patterns that contribute to a more comprehensive understanding of online professional networking dynamics. This would also help enhance the external validity of the findings and provide insights into the broader implications of visual representation in virtual professional spaces.

References

Gunaydin, G., Selcuk, E., & Zayas, V. (2017). Impressions Based on a Portrait Predict, 1-Month Later, Impressions Following a Live Interaction. Social Psychological and Personality Science, 8(1), 36-44. https://doi.org/10.1177/1948550616662123

Hunt, Matthew. "Do Linkedin Profile Photos Matter?" *LinkedIn*, 7 Dec. 2021, www.linkedin.com/pulse/do-linkedin-profile-photos-matter-matthew-hunt/.

Quantity Improvement Solutions. "LinkedIn Profile Photo." *LinkedIn*, 20 Sept. 2021, www.linkedin.com/pulse/linkedin-profile-photo-quantity-improvement-solutions/.

Wargo, Eric. "How Many Seconds to a First Impression? ." *Association for Psychological Science*, 1 July 2006, www.psychologicalscience.org/observer/how-many-seconds-to-a-first-impression.