Gender-Likability IAT: Implicit Likability of Men and Women in Professional Settings

Anke Hao

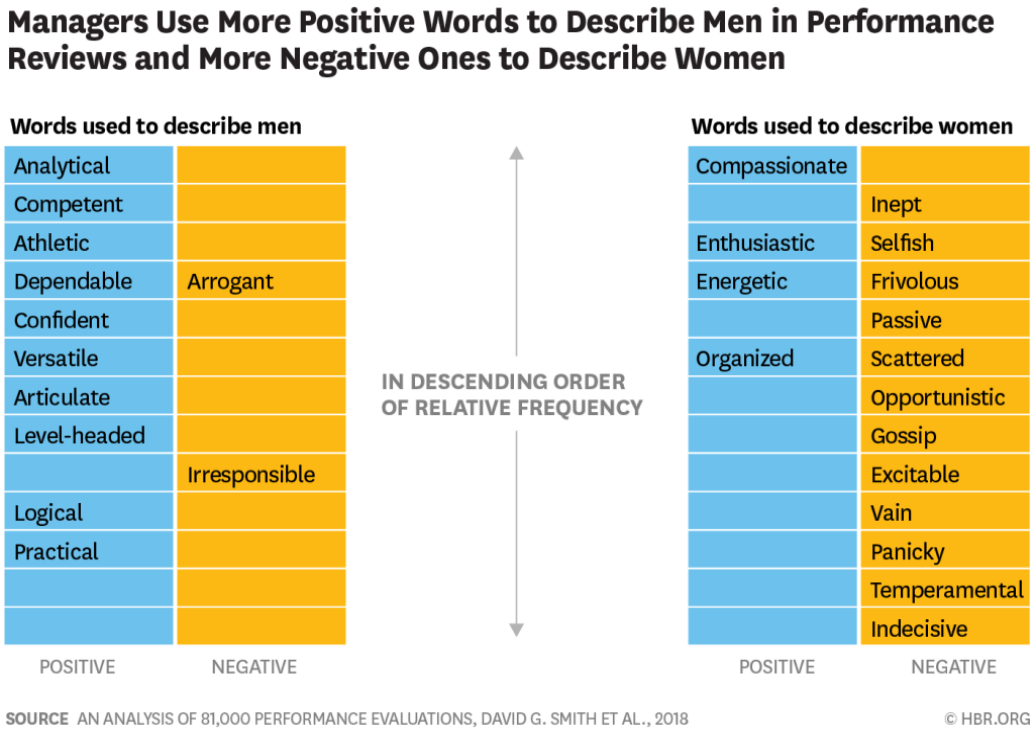
Big Data in the Psychological Sciences

Dr. Bainbridge

August 2nd, 2021

Introduction

Multiple studies have revealed that women have been described in markedly different terms than men in professional settings. In a previous analysis of 81,000 performance reviews in professional work settings, the majority of the most commonly used descriptors for men were positive, while the majority of the descriptors for women were negative (Smith, Rosenstein, & Nikolov, 2021). The top descriptors used are shown in the figure below:



The implications of these patterns are troubling. The top positive descriptors used for men (e.g. Analytical, Competent) are arguably more useful to a company than the top positive descriptors used for women (e.g. Compassionate). In addition, the second most frequently used descriptor for women, "inept," is a warning flag that an employee is unable to do her job—and therefore a company may believe that a decision to fire her would be justified.

Beyond the patterns found in performance reviews, women were also twice as likely to be branded bossy in the workplace, and "bossy women coworkers are seen as more unpopular and less successful compared to bossy men coworkers." In addition, "bossy" women were seen as less promotable compared to "bossy" men (*Ban Bossy: What's Gender Got to Do with It? | CCL*, 2020).

Both studies indicate that women are more likely to be viewed in ways that would have a negative impact on their professional success. What these results do not demonstrate, however, is if this is a result of implicit bias. As such, this study aims to see if there is a demonstrable implicit bias in favor of linking male-associated words with positive characteristics, and female-associated words with negative characteristics. The experiment is an Implicit Association Test that tests for implicit bias between associations of gender and positive or negative characteristics.

What is the Implicit Association Test?

The Implicit Association Test (IAT) measures the "strength of associations between concepts (e.g. black people, gay people) and evaluations (good, bad) or stereotypes (e.g. athletic, clumsy)" (*About the IAT*, 2011). Participants sort stimuli into categories by pressing the corresponding key as fast as they can (usually "e" for left, "i" for right), while their reaction times are recorded. Some of these stimuli would require participants to utilize the top categories to make judgements (e.g. "White" or "Black"), while others would require them to utilize the bottom categories to make judgements (e.g. "Good" or "Bad"). If their reaction time for one particular type of association (e.g. White/Good and Black/Bad for the Race IAT) is faster than the other possible association (e.g. Black/Good and White/Bad), the experiment will conclude that they "implicitly" prefer the association they reacted faster to, as opposed to the association they reacted slower to.

IAT and Big Data

Multiple IAT tests are currently hosted by Project Implicit and open to the public. Anyone interested could take a test in any of several categories listed, including the more well-known Race/"Black-White" IAT, as well as others including the Weight IAT and the Sexuality IAT.

Project Implicit has collected data across the US and internationally, and has made this data anonymized and publicly available. By definition, this would be considered "wide" data, which may be generalizable to the larger population due to the sheer breadth of participants

sampled, but which may not give extremely in-depth insights into any single participant in particular.

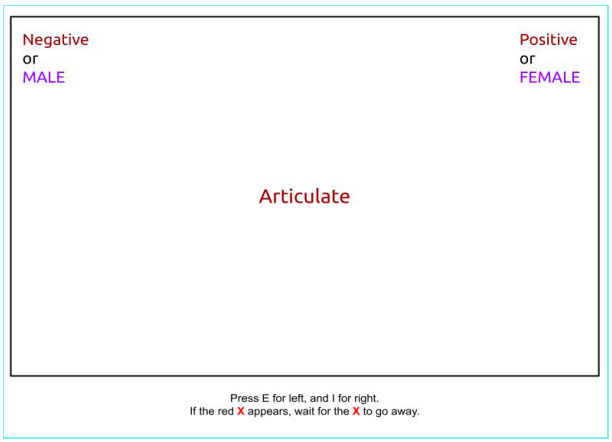
Methodologies

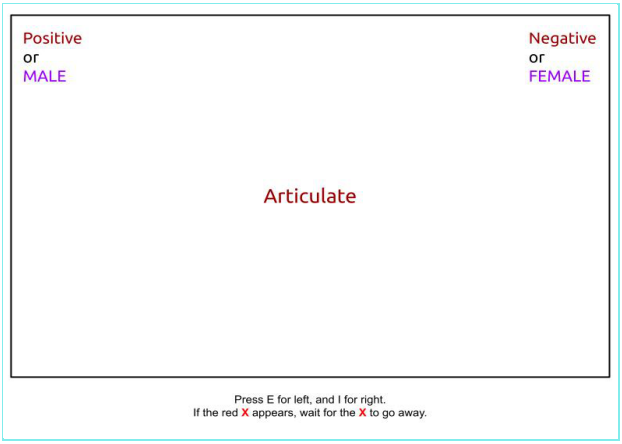
Experiment Overview

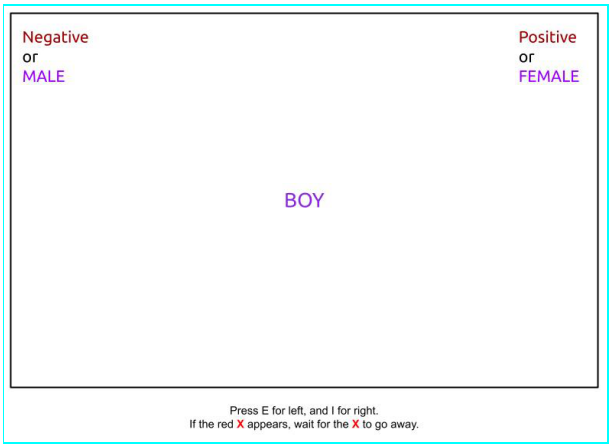
The study was a Psytoolkit experiment that was coded and embedded into a survey that asked preliminary questions about gender, age, and ethnicity. The Psytoolkit experiment recorded reaction times of two different conditions:

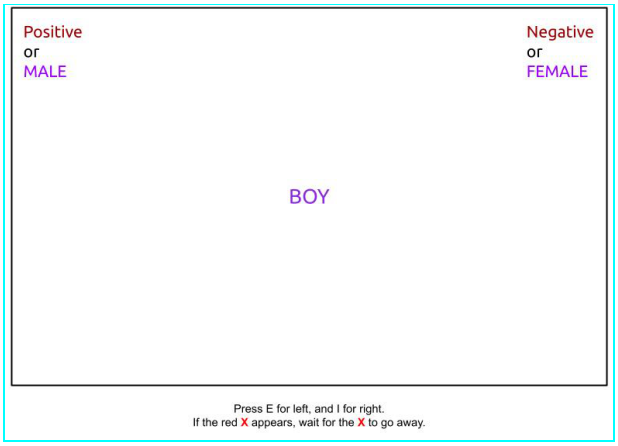
- Linking Negative characteristics with Male-associated nouns and Positive characteristics with Female-associated nouns (NMPF).
- Linking Positive characteristics with Male-associated nouns and Negative Characteristics with Female-associated nouns (PMNF).

Examples of the two conditions are shown below:









In the example on the top left, the word in the center ("Articulate") would need to be categorized into the Positive/Female condition, as "Articulate" is a positive characteristic, while it would be categorized into the Positive/Male condition for the example on the top right. Similarly, the word "Boy" on the bottom left would be categorized into the Negative/Male condition and, on the bottom right, into the Positive/Male condition. Male and Female-associated words used in the experiment were derived from a version of the Gender-Career IAT (*Gender-Career IAT*, n.d.), and a majority of the Negative and Positive-associated characteristics used were derived from an analysis of the most common attributes assigned to men and women in performance evaluations at work (Smith, Rosenstein, & Nikolov, 2021).

Experiment Design Considerations

In order to account for the effect of "practice" in categorizing words, the participants were randomly assigned an order in which they were tested on the two conditions, referred to as a block order. The two random block orders were PMNF first/NMPF second and NMPF first/PMNF second. Regardless of which block order they were assigned, participants got to "practice" more for their second condition by getting more training trials in order to counteract their previously learned association. As an example, if a participant was assigned the PMNF first/NMPF second block order, they would get 5 training trials for the PMNF condition prior to formal trials with data collection, while they would get 10 training trials for the NMPF condition.

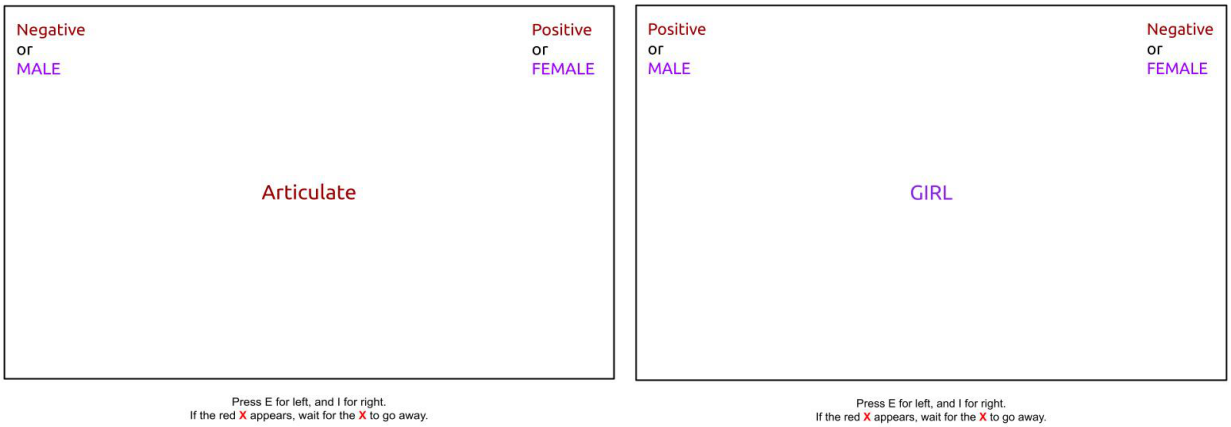
Experiment Design

Participants each went through 7 "blocks," or trials, with 4 blocks total in which data was collected and with 3 blocks as training. An overview of the 7 blocks is listed below:

- **Block 1:** 5 practice rounds for "positive" vs "negative" characteristics
- **Block 2:** 5 practice rounds for "MALE" vs "FEMALE" associated words
- **Blocks 3 & 4:** formal data collection rounds, 10 rounds each (e.g. PMNF)
- **Block 5:** "Positive" and "negative" get switched; 10 practice rounds
- **Blocks 6 & 7:** formal data collection rounds, 10 rounds each (e.g. NMPF)

Each participant experienced both the PMNF and the NMPF conditions, although the order in which they encountered the conditions were randomized. Their task was to press either

the “e” or the “i” key as fast as they could in order to categorize the word they saw in the center of the screen into either the left or right side, depending on the nature of the word. As an example, for the conditions below, the participant would have tried to press the “i” key as fast as possible in order to categorize the word “Articulate” as a Positive characteristic, and “Girl” as a Female-associated word.



Participant Demographics

A total of 369 participants were recruited from Amazon Mechanical Turk to take part in the experiment and were each paid 30 cents for participation in a 2-3 minute experiment. 68 were excluded from final data analysis after data collection due to: 1) unfinished survey entries (n = 37), 2) incorrect answers to the attention check (n = 4) , 3) having an error rate higher than 40% in either block order or having a total error rate higher than 60% (n = 27). This resulted in a total of 301 participants for data analysis (59.5% female).

Table 1: Ethnicity

White	Asian/Pacific Islander	Black/African American	Hispanic/ Latino	Mixed	Other
74.8%	12.6%	7.0%	4.0%	0.3%	1.3%

Table 2: Age

18-24	25-34	35-44	45-54	55-64	65-74	>75
8.3%	29.9%	23.6%	20.3%	10.3%	7.0%	0.7%

Table 3: Gender

Female	Male

59.5%	40.5%

Comparison to U.S. Demographics

Data from the U.S. Census (*U.S. Census Bureau QuickFacts: United States*, 2019) shows that the participants for IAT Gender-Likability are overrepresented in White (60.1% in the U.S.) and Asian/Pacific Islander (5.9% in the U.S.) ethnicities, and under-represented in the Black/African American (13.4% in the U.S.) and Hispanic/Latino (18.5% in the U.S.) ethnicities. Furthermore, there was an overrepresentation of females in the study (59.5% in the study compared to 50.8% in the U.S.). Although segmented age demographics are not available in the U.S. Census (i.e. breakdowns in age groups between ages 18-64), younger participants seem slightly overrepresented in the study, with 38.2% under 35 years old and 61.8% under 45 years old. The majority of the age range of 18-44 is comprised of the Gen Z and Millennial generations.

Results

There was no significant difference in average reaction times with the PMNF versus the NMPF order, regardless of assigned block order. Combining the two block orders resulted in the mean and median response times and error rate (e.g. pressing "i" instead of "e") as follows:

Table 4: Both Block Orders Combined	
Mean (NMPF)	**Mean (PMNF)**
1065.6 ms	1089.0 ms
Median (NMPF)	**Median (PMNF)**
917.8 ms	954.3 ms
Error rate (NMPF)	**Error rate (PMNF)**
6.1%	6.6%

The results were then segmented into the specific block order in which the participants encountered the experiment. Block order NMPF/PMNF, as displayed on the left side of the table below, indicates that the participant first went through rounds where Negative characteristics were linked with Male-associated words, and Positive characteristics were linked with Female-associated words, then went through rounds where Positive characteristics were linked

with Male-associated words, and Negative characteristics were linked with Female-associated words in the later half. The block order PMNF/NMPF on the right side of Table 5 indicates the opposite order.

Table 5: Separate Block Orders			
Block Order NMPF/PMNF		**Block Order PMNF/NMPF**	
Mean (NMPF)	**Mean (PMNF)**	**Mean (NMPF)**	**Mean (PMNF)**
1061.7 ms	1073.9 ms	1069.5 ms	1104.1 ms
Median (NMPF)	**Median (PMNF)**	**Median (NMPF)**	**Median (PMNF)**
928.1 ms	949.9 ms	907.6 ms	958.8 ms
Error rate (NMPF)	**Error rate (PMNF)**	**Error rate (NMPF)**	**Error rate (PMNF)**
6.0%	6.8%	6.2%	6.3%

The response time for the NMPF condition was faster than the PMNF response time, and the error rate for the NMPF condition was lower than the PMNF condition. This implies that participants were both faster and more accurate when linking negative words with male-associated words and positive words with female-associated words.

To further explore the effect of gender on participant results, the results were then segmented by participant gender. Almost all of the participants (n = 298) identified as female or male. There were two participants who identified as nonbinary and one who identified as “other.” As their results would not be representative of their population, they were excluded from the following table.

Table 6: Results by Gender			
Female		**Male**	
Mean (NMPF)	**Mean (PMNF)**	**Mean (NMPF)**	**Mean (PMNF)**
1061.7 ms	1073.9 ms	1069.5 ms	1104.1 ms
Median (NMPF)	**Median (PMNF)**	**Median (NMPF)**	**Median (PMNF)**
928.1 ms	949.9 ms	907.6 ms	958.8 ms
Error rate (NMPF)	**Error rate (PMNF)**	**Error rate (NMPF)**	**Error rate (PMNF)**
5.4%	7.7%	7.0%	5.7%

The results in Table 6 show that female participants have a faster reaction time and a lower error rate in NMPF conditions, as opposed to in PMNF conditions. The opposite is true for male participants. This means that female and male participants 1) respond faster and 2) more accurately to associating positive words with words linked to their own gender instead of the opposite gender.

However, the results were not significant due to a high standard deviation in all cases. Table 7 displays the standard deviation for the aggregated results. The standard deviations for the results when separated by block order and by gender can be found in the IAT results data.

Table 7: σ for Aggregated Results

Mean (NMPF)	**Mean (PMNF)**
521.7 ms	411.2 ms
Median (NMPF)	**Median (PMNF)**
327.2 ms	337.7 ms
Error rate (NMPF)	**Error rate (PMNF)**
7.7%	7.0%

Although the results are not significant, the average median response time for females participating in the experiment was faster for the NMPF condition (868.8 ms) than for the PMNF condition (895.7 ms), while the opposite was true for male participants' average median response times (877.3 ms for the NMPF condition, 868.2 ms for the PMNF condition). Overall, it seems that both females and males prefer to associate their own gender with positive words over negative words.

Discussion

As the difference in two means was not significant, the average reaction times of the participants do not conclusively favor one condition over the other (PMNF vs NMPF). There are several factors that may contribute to this, including a high proportion of female participants, a younger

demographic among the participants, a small sample size, and non-standardized testing conditions.

High Proportion of Female Participants

The proportion of female participants was 59.5% in the study, which over-represents the 50.8% in the U.S. demographic. While further research would be required, a potential contributing factor to how participants did not have an implicit preference for one condition over another may be due to female overrepresentation and potentially a subsequent effect in favor of Positive/Female associations, as seen in the results where females had a faster response time to Negative/Male-Positive/Female conditions.

Younger Demographic

Over half of the participants were Gen Z and Millennials, and both generations are more known for more liberal and progressive viewpoints (Parker et al., 2020) (*The Generation Gap in American Politics*, 2019). This may be reflected in the sample if explicit attitudes have begun translating into implicit attitudes about how to view behaviors and personalities in either gender.

Small Sample Size

Due to budget constraints, the data collected in the current study constitutes a total of 301 data points. With more data, the results may be more conclusive and potentially more representative, as the confidence level would be higher.

Non-standardized Testing Conditions

The participants took part in the study with their own computers, keyboards, and environments. Their computers and keyboards may have differences in quality that impeded or helped their reaction times, and some participants may have been in a noisy or otherwise distracting environment. These differences would make a big impact on reaction time results, especially as they were measured in milliseconds. If we could have collected responses in a standardized environment with standardized equipment, we would have mitigated this effect.

Future Directions

Future experiments with the Gender-Likability IAT, beyond collecting more data, may include new ways of segmenting data, surveying other populations, and enhancing the stimuli.

New Segmentation

With more participants and more funding to allow for longer preliminary surveys on demographics, new questions on the preliminary surveys may ask about participants' geographic locations (e.g. state, country), political affiliation, education, socioeconomic status, religion, and social advocacy involvement. With a larger dataset, these questions would allow participants to be segmented into categories for further analysis in order to see if there would be significant differences between different cohorts (e.g. if a particular state or region of the U.S. would respond differently than another state or region).

New Populations

All the participants surveyed identified overhelmingly as female or male, so new populations that could potentially be surveyed include non-binary or gender-queer individuals. Furthermore, since the study took place on Amazon Mechanical Turk, participants had to be 18 years or older. However, another variation of this study could survey children from elementary school to high school, which could potentially investigate if there are changes in implicit preferences over time while children are still in the development phase, and if children's perspectives are different from adults'. International populations could also be surveyed, if the stimuli are adjusted to reflect languages that the participants are fluent in.

New Stimuli Enhancements

In order to expand reach to international populations, the stimuli would need to be translated into different languages. Visual stimuli could also be introduced (e.g. a female face), although these stimuli would first need to be cross-checked by crowdsourcing opinions on them to make sure confounding variables like the attractiveness of the face or a specific facial expression would affect the participants' judgement of "likability" as little as possible. Another option would be to introduce more "gendered" likability characteristics to introduce more variation in stimuli.

Works Cited

About the IAT. (2011). Project Implicit: About the IAT. https://implicit.harvard.edu/implicit/iatdetails.html

Ban Bossy: What's Gender Got to Do with It? | CCL. (2020, November 18). Center for Creative Leadership. https://www.ccl.org/articles/white-papers/bossy-whats-gender-got-to-do-with-it/

Gender-Career IAT. (n.d.). Retrieved March 04, 2021, from https://implicit.harvard.edu/implicit/user/agg/blindspot/indexgc.htm

Parker, K., Graf, N., & Igielnik, R. (2020, May 30). *Generation Z Looks a Lot Like Millennials on Key Social and Political Issues*. Pew Research Center's Social & Demographic Trends Project. https://www.pewresearch.org/social-trends/2019/01/17/generation-z-looks-a-lot-like-millennials-on-key-social-and-political-issues/

Smith, D. G., Rosenstein, J. E., & Nikolov, M. C. (2021, January 20). *The different words we use to describe male and female leaders.* Retrieved March 04, 2021, from https://hbr.org/2018/05/the-different-words-we-use-to-describe-male-and-female-leaders

Stoet, G. (2010). *PsyToolkit - A software package for programming psychological experiments using Linux.* Behavior Research Methods, 42(4), 1096-1104.

Stoet, G. (2017). *PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments.* Teaching of Psychology, 44(1), 24-31.

The Generation Gap in American Politics. (2019, December 31). Pew Research Center - U.S. Politics & Policy. https://www.pewresearch.org/politics/2018/03/01/the-generation-gap-in-american-politics

U.S. Census Bureau QuickFacts: United States. (2019). Census Bureau QuickFacts. https://www.census.gov/quickfacts/fact/table/US/PST045219