

# The implicit association test – A method to create awareness on diversity and inclusion?



*In collaboration with*



Master Thesis Behavioral Economics

Name student: Isabelle Patist

Student ID number: 510317

Supervisor: Aurélien Baillon

Second assessor: Sophie van der Zee

Microsoft supervisor: Lisanne Brons

Date final version: 27-08-2019

## **Abstract**

The Implicit Association Test (IAT) has become an increasingly popular method to make people aware of their unconscious biases, especially on diversity and inclusion (D&I) issues within companies. However, the literature researching whether this is an effective method, is limited. The aim of this research is therefore, to create a design for companies to assess whether the IAT can be used to create awareness for D&I, within their organization. The research question central in this study is: Does completing an IAT on discrimination have an effect on a person's awareness and behavior towards D&I? A newly developed design was tested as a pilot at Microsoft the Netherlands. The data generated from this design is then analyzed in three ways: overall analysis, single question analyses and principal factor analysis. In this pilot experiment, all three analyses methods showed no significant difference in awareness related to D&I between the subjects that took the IAT, compared to the subjects that did not. Nevertheless, the results are subject to a low sample size and sample variety. As a result, the contribution of this paper is the design of the conducted pilot experiment, and a proposed quantitative analysis for the gathered data.

**Key words:** Diversity, Inclusion, Awareness, Implicit Association Test (IAT), Unconscious bias, Sensitive questions.

**Table of Content**

- 1. Introduction ..... 3
- 2. Literature review ..... 6
  - 2.1. Diversity and inclusion ..... 6
  - 2.2. The Implicit Association Test..... 7
  - 2.3. Sensitive question methodology ..... 11
- 3. Methodology ..... 12
  - 3.1. Experimental design ..... 12
  - 3.2. Data ..... 18
  - 3.3. Hypotheses & Models..... 22
- 4. Results ..... 24
  - 4.1. Overall analysis ..... 24
  - 4.2. Single question analyses ..... 26
  - 4.3. Principal factor analysis..... 29
- 5. Discussion, limitations and future research ..... 31
  - 5.1. Discussion ..... 31
  - 5.2. Limitations and additional considerations ..... 31
- 6. Conclusion ..... 32
- 7. References ..... 34
- 8. Appendix ..... 37

## 1. Introduction

On May 17<sup>th</sup> 2019, a Dutch newspaper called the NRC (Vriesinga, 2019) published an article about the upcoming trend in Dutch businesses to invest more in diversity and inclusion, such as in training employees about the benefits of a more diverse and inclusive workforce. In this article, data from research of TNO and CBS suggests that while 35% of employed males had a supervisory role in the Netherlands in 2018, the percentage among females was much lower (18%). To address this inequality, the Dutch government has since 2013 enforced the Act on Management and Supervision on large companies<sup>1</sup> (Rijksoverheid, 2017). The act entails that these companies should strive<sup>2</sup> for a balanced distribution of at least 30% males and 30% females in their management boards and supervisory boards. However, this target figure has to this day not been met for either the management boards or supervisory boards. In 2016, the Dutch Companies Monitor Women at the Top (Bedrijvenmonitor Topvrouwen) found that the average share of women in management boards is 10.7%, whereas the average share of women in supervisory boards is 15%. Moreover, 55% of the companies that did not comply with this act, did not mention this in their annual report (Pouwels & Henderikse, 2017).

Prioritizing diversity and inclusion (D&I) is an increasingly important topic, not only in the Netherlands, and involves more than just gender equality. Diversity is a broad concept that is defined by the Ferris State University as “the range of human differences, including race, gender, ethnicity, sexual orientation, age, social class, physical ability or attributes, religion, nationality, and political beliefs” (Ferris State University, 07-06-2019). Inclusion refers to the behavior in which no one is discriminated against because of his or her differences. In order to expand D&I, governments and firms from all over the world invest large sums of money. But what are the benefits associated with D&I?

One of the arguments that has been used to increase D&I are the ethical drivers such as fairness for everyone and lower discrimination. Moreover, studies found that from an economic and psychological perspective, several benefits are associated with a higher rate of diversity in the workplace. Herring (2009) argues that diversity in race is linked with increased sales revenue, market share, relative profits and a higher number of customers. Herring also found that sales revenue, customer base and relative profits increase with gender diversity. Also, a study by the Boston Consulting Group (2017) suggests that a higher level of innovation and

---

<sup>1</sup> Companies with assets over €17,5 million, net turnover of €35 million and a minimum of 250 employees.

<sup>2</sup> If a company does not comply with this act, this company must explain why not and how it plans to comply with the act in the future.

improved financial performance is generated through an increase in diversity of leadership teams. In addition, a study by Roberge and van Dick (2010) found that diversity leads to a higher group performance when a larger variety of social psychological mechanisms is used.

Once D&I becomes a priority for a firm, it must develop a strategy to achieve this goal. The choice of an effective diversity and inclusion strategy depends on a number of factors, such as firm size, firm and country culture, current diversity rate, methods the firm uses to measure D&I, the firm's industry, etc. In the past decade, several measures have been used to boost D&I, but due to variety of factors that influence the choice, there is no one-size-fits-all solution. Targets for recruiters, trainings on the importance of diversity for employees (Tipper, 2004), changing job descriptions and anonymizing CV's (Hiscox et al., 2017) are all examples of measures taken by firms and governments to boost D&I. However, not all of them have led to the expected outcomes. In Australia a trial was conducted by the behavioral economics team of the Australian government (BETA) in which the impact of de-identifying applications for senior positions in the Australian Public Service has been analyzed. "The aim was to assess whether de-identifying applications would, by eliminating the effects of explicit or implicit bias, help promote gender equality and diversity in hiring at senior levels." (Hiscox et al., 2017). However, Hiscox (2017) found that this measure led to an undesirable outcome. Fewer women were hired through the policy because the public employees engaged initially in positive (not negative) discrimination towards female and minority candidates.

Furthermore, to measure the effects of the implemented D&I strategies, firms use a variety of metrics. For diversity in gender, nationality, physical ability and age group, metrics such as employee statistics give a clear overview. However, other human differences like sexual orientation and political beliefs are more difficult to measure. Especially since these are not immediately visible or are allowed to be registered. Additionally, inclusion is intangible and requires therefore a different method to measure. Within Microsoft, every year a companywide survey is distributed. In this survey, several questions concerning their employees' opinion about the company's culture are posed. This is one explicit way of measuring the level of inclusion of an organization. Another metric that is used within companies is the Implicit Association Test (IAT). The IAT is developed in 1998 by researchers of Harvard, the University of Washington and the University of Virginia called "project implicit". The objective of this test is to measure implicit attitudes (Greenwald, McGhee & Schwartz, 1998). Implicit attitudes are certain actions or judgements of a person as a response to an automatic evaluation, without this person's realization of this (Greenwald and Banaji, 1995). A type of an implicit attitude is

the unconscious bias<sup>3</sup>. Having an unconscious bias often leads to discrimination of which a person is not fully aware (Lee, 2005). The original IATs developed by the “project implicit” team are mainly focused on measuring discrimination, but similar tests are nowadays developed for a range of topics. Jajodia and Earleywine (2003) have for example used an IAT to measure alcohol usage expectancies and Maison et al. (2001) to measure consumer attitudes.

Besides the measurability component of the IAT, firms are increasingly using the test to create awareness within their company. One of these companies is Starbucks. On May 4<sup>th</sup> 2018, Starbucks Canada tweeted that they would close their stores on the afternoon of June 11<sup>th</sup> 2018 to provide all their employees with an ‘Implicit bias & conscious inclusion training’ (Starbucks Canada, 2018). Nevertheless, the effectiveness of these trainings is difficult to test, since there is a variety of factors that could influence the outcomes. Atewologun et al. (2018) analyzed evidence for the effectiveness of unconscious bias trainings. They found in their study that by using the IAT in combination with educating employees on unconscious bias theory, it will likely create awareness and reduce implicit bias.

However, the literature on this is limited. There is not yet a clear methodology developed to test the effectiveness of the IAT on raising awareness, and influencing behavior within a company. The aim of this paper is therefore to fill this gap in literature. Specifically, to introduce an experimental design that tests whether using the IAT is effective as a method to raise awareness and influence behavior within a company. This experiment is implemented as a pilot at Microsoft the Netherlands (Microsoft NL). The research question is therefore:

*Does completing an IAT on discrimination have an effect on a person’s awareness and behavior towards D&I?*

This paper consists of the following chapters. Chapter two presents the literature review on diversity and inclusion, the Implicit Association Test and the methodology behind sensitive questions, followed by chapter three in which the methodology of the conducted experiment is described in detail. Chapter four presents the analyzed results of the experiment after which in chapter five follows a discussion of these results. Finally, in chapter six the limitations of this research are discussed and how future research can build on this paper.

---

<sup>3</sup> Unconscious bias is defined as “a prejudice or unsupported judgment in favor or against one thing, person, or group compared to another, in a way that is usually considered unfair” (Vanderbilt.edu, 2019).

## 2. Literature review

In this section, literature is presented on diversity and inclusion, the Implicit Association Test and sensitive question methodology.

### *2.1. Diversity and inclusion*

As mentioned in the introduction, a more diverse and inclusive workforce creates various benefits for organizations. However, firms and governments report that the transition towards greater D&I goes slower than they anticipated (Hunt, Yee, Prince, 2018).

From a behavioral economics perspective there are several biases and heuristics that arise when implementing a culture transformation. These biases and heuristics can limit the effect of the taken measures to improve D&I (Samson, 2018). First, the organization must come up with a goal to improve D&I, and then define a plan to reach this goal. However, the created plan to reach this goal is not always suitable. One of the reasons can be the presence of overconfidence bias during the development of the plan. Fischhoff, Slovic and Lichtenstein (1977) found in their research that overconfidence occurs when the decision makers are too optimistic in their primary evaluation of the situation. In addition, Busenitz and Barney (1997) argue that this effect is often combined with a slow response in incorporating additional information, due to the overconfidence in the beginning.

Moreover, Samuelson and Zeckhauser (1988) found that when decision makers are confronted with new options for the course of the company, the decision makers often stay with the status quo alternative. This is also referred to as the status quo bias. In addition, the status quo bias can also affect the behavior of the employees. In the study of Furst and Cable (2008), the resistance of employees towards organizational change has been investigated. They argue that most of the time, efforts of decision makers to implement an organizational change fail. Additional, Furst and Cable (2008) mention that decision makers should make it clear to their employees that status quo behavior is unacceptable.

Herding behavior (Samson, 2018) is another behavior that can influence the effect of a culture transformation in large organizations. Samson mentions that herding occurs when people mirror each other's behavior instead of basing their decisions on their own evaluation of information. This can have a positive or negative impact on the imposed change plan, dependent on whether the behavior of people who are in favor of the D&I change is mirrored, or that of those against. Moreover, Cialdini (2005) mentions consensus as one of the six principles of persuasion. He argues that by pointing out what others have done in line with the

required change before, other people are more likely to follow. Organizations can use this in their advantage.

The ‘unconscious’ in the unconscious bias is another argument for a reduced effect of a measure taken to improve D&I (Lee, 2005). Without being aware of sometimes adopting discriminating behavior, he or she might not see the need for change to increase inclusion and therefore not contribute to the organization’s D&I goal.

Once employees agreed to the plan to improve D&I, other biases and heuristics can arise. One of which is the intention-behavior gap. Sheeran and Webb (2016) studied the situation in which a person’s behavior does not match his or her intention, leading to an intention-behavior gap. They found that only one-half of the occurrences, intentions became translated into action.

Lastly, when decision makers implement several initiatives to increase their organization’s D&I, their employees can suffer from ego depletion. Baumeister et al (1998) describe ego depletion as that a person’s self-regulation is a limited resource. In other words, that once a person has exercised a self-regulating task, the ability to complete future self-regulating tasks becomes more difficult for him or her. In this case, this could lead to D&I ‘fatigue’.

## ***2.2. The Implicit Association Test***

As mentioned before, the IATs created by the “project implicit team” measure a person’s implicit attitudes and beliefs that he or she may be too ashamed of to report. The IAT also analyzes the intensity of this person’s implicit attitudes, and whether this person has a stronger implicit automatic association towards a certain topic, compared to another.

Examples of IAT’s that are available online at the creator’s website [implicit.harvard.edu](http://implicit.harvard.edu) are: Weight, Age, Skin-tone, Disability and Gender & Career. The following table 1 illustrates how the IAT Gender and Career works.

Series	1		2		3		4		5	
Task description	Initial target-concept discrimination		Associated attribute discrimination		Initial combined task		Reversed target-concept discrimination		Reversed combined task	
Category	Male	Female	Career	Family	Male & Career	Female & Family	Female	Male	Female & Career	Male & Family
Sample stimuli	Ben Paul Daniel John Jeffrey	Rebecca Michelle Emily Julia Anna	Corporation Salary Office Professional Management Business	Marriage Parents Relatives Family Home Children	Examp.: Ben – Salary  John - Office	Examp.: Emily – Home  Julia - Parents	Rebecca Michelle Emily Julia Anna	Ben Paul Daniel John Jeffrey	Examp.: Emily – Salary  Julia - Office	Examp.: Ben – Home  John - Parents

Table 1. The five rounds of the IAT Gender Career.

Table 1 describes the five series of the IAT process for gender and career. Before the first series starts, the categories male, female, career and family are displayed with their sample stimuli (Greenwald, McGhee & Schwartz, 1998). This way the participant knows which stimuli belongs to which category, and which match is considered a correct answer. Then in the first series, initial target concept discrimination, the subject is asked to identify names with gender. The subject assigns these names on the computer to the correct category by a response by the left hand on the “e” key or right hand on the “i” key. In the second round, the subject is asked to perform a similar task, namely assigning concepts to either career or family. For example, if the category career was on the left and the word “office” appeared on the screen, the subject must press the “e” key. The third round combines the categories and word groups. Meaning, when on the left the two categories Male & Family (and on the right side Female & Career) are displayed<sup>4</sup>, the subject has to assign either male names or family concepts to the left by pressing the “e” key (and female names and career concepts to the right by pressing the “i” key). In the appendix the figures A and B show how this looks like when making the test. The fourth round is similar to the first round, but this time the genders have switched sides. Also, the last round, number five, is similar to the third round where both categories are combined. However, here the genders are paired with the other group of either career or family (see figures A and B in the appendix).

<sup>4</sup> Because of the randomization of the order, this round could also be Female & Family and Male & Career.

During this test, duration and mistakes when allocating the words into groups are monitored and the results of the rounds are compared to evaluate whether the participant finds it unconsciously easier to make associations between either male and career (female and family) or female and career (male and family). The result that the subject receives can deviate between the following seven outcomes as shown below in figure 1.

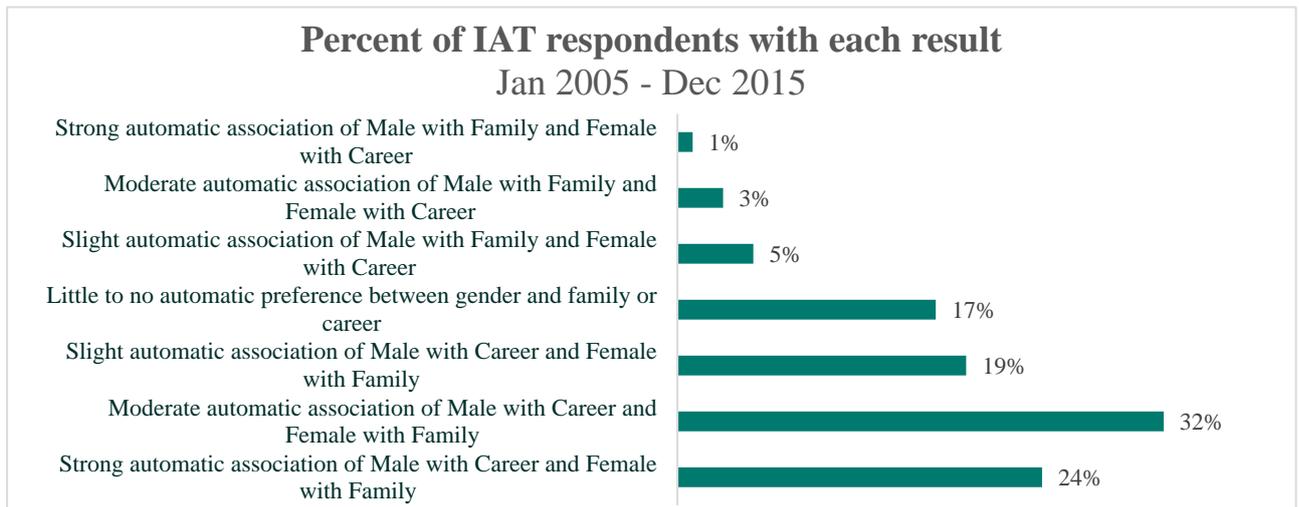


Figure 1. The distribution of 846 020 IAT scores for the Gender-Career task completed between January 2005 and December 2015 (Project Implicit by Harvard, 2016). Visible here is that the highest percentage (32%) of people have a moderate automatic association of Male with Career and Female with Family, whereas the smallest percentage (1%) consist of people that have a strong automatic association of Male with Family and Female with Career.

Nowadays, the IAT is often used as a measurement tool for implicit discriminating behavior, especially since studies show that this is a preferred measuring method compared to self-reported methods (Greenwald, Banaji & Nosek, 2015). For example, Tourangeau and Yan (2007) are in favor of the IAT approach, because of the lower chance of misreporting by the subjects on the sensitive topics tested through the IAT. Due to the non-behavioral task of matching words, subjects have less power to consciously misreport their true preference compared to explicit measures such as self-reporting. In line with this are the findings of Frantz et al (2004), who suggest that the IAT is valuable in investigations of sensitive topics such as racial attitude by avoiding the social desirability issues. Moreover, they argue that the IAT is easy to use and creates robust results, which makes it attractive to use in research.

On the contrary, previous literature has also found results not favoring the IAT approach to measure or predict behavior. Karpinski and Hilton (2001) conducted three types of studies where the relationship between the IAT and explicit attitudes were examined. In the first study they compared the implicit attitudes measured by the IAT with explicit attitudes. They found that there is no correlation between the IAT results and explicit attitudes, regardless whether

the two measures show the same preferences of the participants. In addition, Karpanski and Hilton found in their second study that the IAT failed to predict behavior in the choice between a candy bar or an apple. The authors do however stress that this result only applies to this specific choice, and that the IAT may be a good predictor of behavior in other situations. Lastly, in their third study they again found no correlations between explicit attitudes and the implicit attitudes found by the IAT. The authors conclude from these three studies that there is strong evidence for the independence of the IAT and explicit attitude measures. Additionally, the study shows that the IAT scores do not so much mirror the extent to which the person approves the measuring associations. However, it does show the associations that the participants have been exposed to in their environments.

Olson and Fazio (2004) argue that using the traditional IAT as a measurement, does not account for extra-personal associations. Extra personal associations are described as associations that are accessible in the memory, but are considered irrelevant since they do not become activated when the subjects interacts with the attitude object tested by the IAT. In other words, the IAT also includes the difficulty that a participant experiences with connecting certain words to given categories. Associations that the participant makes that are unrelated to the participant's actual evaluation of this category can therefore influence the outcome of the IAT. Olson and Fazio conducted four experiments in which a 'personalized IAT' was compared to the traditional IAT. They found that by small adjustments to the attribute category labels and attribute items the effect of extra personal associations is reduced. The adjustments they made in the attribute category labels was from pleasant / unpleasant to I like / I do not like. Moreover, they used words which had little social consensus among the participants and excluded the error feedback to personalize the IAT.

These studies (Greenwald, Banaji & Nosek, 2015. Tourangeau & Yan, 2007. Frantz et al., 2004. Karpinski and Hilton, 2001. Olson & Fazio, 2004) focus on the measurability component of the IAT on behavior. However, the team that developed the IAT stresses that the IAT should only be used as an educational tool to develop awareness of implicit preferences and possible stereotypes. The use for significant decision-making such as choosing between job applicants is strongly advised against (Greenwald, Banaji, Nosek, 1998). Nevertheless, as mentioned before, the literature about the usage of the IAT for an educational purpose is limited.

One of the studies that used the IAT for an educational purpose is Miller et al. (2013). Miller et al. (2013) researched the occurrence of medical students having a weight-related bias. Secondly, Miller et al (2013) studied whether the medical students that had a weight-related

bias, were aware of this. This awareness is determined through evaluating the correlation between the student's IAT scores and their explicit preferences. Miller et al. (2013) found that where one-third of the students had a weight-related bias, 23% of these students were unaware of this. The researchers reported the scores back to the medical students to educate them on the negative effects this can have on the medical students' behavior towards overweight patients.

Green et al. (2007) found in their study that completing the IAT led to a higher awareness. They measured this effect on awareness through the acknowledgement of the participants to a higher vulnerability to unconscious bias, after taking the IAT. However, this result could be subject to a response bias. Meaning, the participants that agreed with a higher experiencing vulnerability, could be influenced by their awareness of being in an experiment, by being socially desirable or other factors. Morris and Ashburn-Nardo (2009) reported a similar finding as Green et al. (2007). Nevertheless, in contrast to Green et al. (2007), Morris and Ashburn-Nardo (2009) performed a dependent-samples t-test. They found that students, that had taken the IAT and participated in classroom discussions about this topic, experienced a significant higher level of awareness of possibly having an unconscious bias themselves or within others. A limitation of this study is however that they could not separate the effect of the IAT and the effect of the classroom discussions. Furthermore, Morris and Ashburn-Nardo (2009) found that the students that took the IAT also performed better at answering questions about the unconscious bias. The outcomes of the research of Atewologun et al. (2018) is in line with these finding. Suggesting that the IAT, in combination with education on unconscious bias, increases awareness on the implicit bias.

### ***2.3. Sensitive question methodology***

Diversity and inclusion are considered sensitive topics. Questions about sensitive topics can lead to several disadvantages for conducting research in this field. One of these disadvantages is a lower unit response rate. Unit response rate is the overall rate of people participating in the experiment or survey (Tourangeau & Yan, 2007). Another effect that Tourangeau & Yan (2007) mention, is a lower rate of people who have agreed to participate, but declined to answer the questions. This is also known as item response rate. To minimize these effects, Tourangeau and Yan state that these sensitive questions should be presented in such a way that the subjects feel comfortable and safe in stating their true preference.

In a previous study, Tourangeau, Rips and Raskini (2000) define three meanings of the concept sensitivity in surveys, namely: 1) intrusiveness, 2) the threat of disclosure and 3) socially undesirability. Intrusiveness in sensitive survey questions indicates that the topic is not part of an everyday conversation; the topics are considered an invasion of someone's privacy. The topics in this paper, diversity and inclusion, are nowadays increasingly openly discussed at work, but since the chance of offending someone is high, not everyone feels as comfortable sharing her or his honest opinion. The second component Tourangeau et al (2000) emphasize is the threat of disclosure. This refers to the concerns of subjects that their answers will be shared with a third party that can lead to negative consequences for them. Social desirability is the last component, meaning the question can draw out answers that are socially unacceptable. Tourangeau and Yan (2007) argue that by adding the option to refuse to answer, the social desirability bias will be decreased. In other words, by having the option to not answer the question, the subjects are less inclined to answer in a way that they think other people will find desirable.

A measure that has been used to test for false reporting is the bogus pipeline (Rasinski et al, 2005). With this method, researchers convince the participants that they can detect lying by using techniques such as polygraph devices or biological tests. Previous studies about smoking and drug behavior have found that when taking saliva or urine tests before asking the respondents about their behavior, the rate of respondents stating their true behavior was higher than in the group of respondents where the biometrics were taken after the questions.

### **3. Methodology**

In this chapter, the pilot experiment that took place at is explained in further detail, followed by a description of the gathered data and the hypotheses that will be tested.

#### ***3.1. Experimental design***

The experiment took place on Tuesday 21<sup>st</sup> and Thursday 23<sup>rd</sup> of May 2019 at noon, both in the same week at the Microsoft NL office. In a closed and spacious room, the subjects were seated away from each other to ensure that they could not see what the other participants were doing. I then gave a short introduction stressing two aspects that the subjects should keep in mind. First, that all the data will be anonymized for everyone, including me. The second aspect is that the purpose of this research is scientific, and it is therefore of importance that they share their real opinion and do not answer based on what is socially desirable. To emphasize the scientific

objective of this experiment, there is a logo of the Erasmus Behavioral Lab presented in the top right corner of the screen throughout the whole experiment. For the reason mentioned before, the option to refuse to answer is also included.

When this introduction was given, I shared an anonymous link with all the subjects (see appendix x for full questionnaire). This link directed the subjects to the Qualtrics environment in which the experiment was programmed. As visible in the figure 2 below and the questionnaire, the subjects were asked three questions. These questions asked the subjects about how important they find liberal arts, how important they find science and what they expect their level of unconscious bias to gender and science or liberal arts to be. After these three questions, the subjects were randomly assigned to either the treatment group or the control group

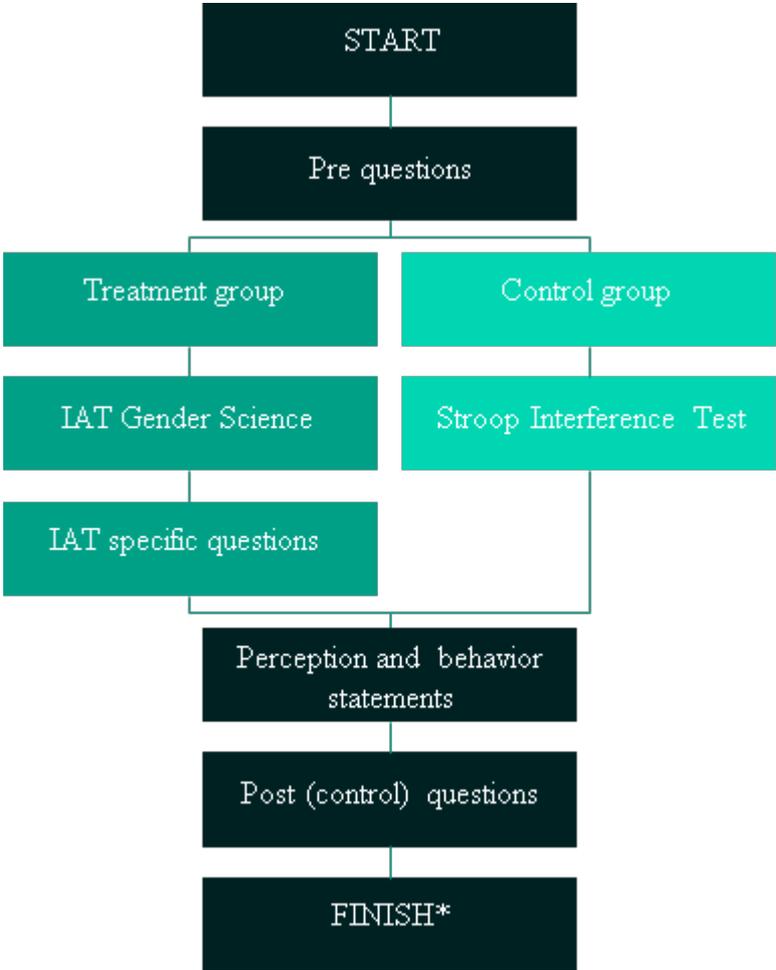


Figure 2. Process flow of the experiment. Through random selection will half of the subject group be allocated in the treatment group, and the other half in the control group. The control group will receive their individual scores after they submitted their responses.

### 3.1.1. Treatment group

The treatment group were presented a screen (figure C in the appendix) in which they were asked to take the IAT Gender Science on the official Project Implicit website ([www.implicit.harvard.edu/implicit/takeatest.html](http://www.implicit.harvard.edu/implicit/takeatest.html)). Ideally, the IAT would be programmed within the same Qualtrics environment as the rest of the experiment, but due to technical difficulties this was not possible. More about this in chapter 5.

The reasoning behind the choice of the IAT Gender Science for this experiment is the type of employees that work at Microsoft and the interest of Microsoft in this specific unconscious bias. Microsoft is known worldwide for the technical products and software they offer, and the majority of its employees have a scientific background. However, Microsoft worldwide suffers not only from an imbalance in gender (27% females to 73% males), but also in race and ethnicity. Microsoft reports that they have been trying to decrease these inequalities for years (Microsoft.com). Understanding the unconscious bias of its employees towards gender and science can help accelerate this growth towards equality.

As described in chapter two, the IAT consists of five rounds. In the following table 2 the rounds of the IAT Gender Science with the used stimuli words are shown. This table is similar to table 1 in chapter two, except for the categories and sample stimuli. The process of clicking left or right in the test and the randomization between subjects in rounds three and five is therefore identical to that explained before.

Series	1		2		3		4		5	
Task description	Initial target-concept discrimination		Associated attribute discrimination		Initial combined task		Reversed target-concept discrimination		Reversed combined task	
Category	Male	Female	Science	Liberal Arts	Male & Liberal Arts	Female & Science	Female	Male	Female & Liberal Arts	Male & Science
Sample stimuli	Man Son Father Boy Uncle Grandpa Husband Male	Mother Wife Aunt Woman Girl Female Grandma Daughter	Astronomy Math Chemistry Physics Biology Geology Engineering	History Arts Humanities English Philosophy Music Literature	Exam.: Boy – History  Husband – Music	Exam.: Wife – Biology  Girl – Geology	Mother Wife Aunt Woman Girl Female Grandma Daughter	Man Son Father Boy Uncle Grandpa Husband Male	Exam.: Wife – History  Girl – Music	Exam.: Boy – Biology  Husband – Geology

Table 2. Conducted IAT Gender-Science with its five rounds.

Once the subjects finished the IAT Gender Science, the website showed their implicit attitude outcomes. They were asked to return to the Qualtrics environment of the experiment and answer two questions only shown to the treatment group. These questions consist of reporting their IAT score and whether their perception of themselves regarding D&I changed. These two questions were not used further in the analysis of this paper, but were presented back to Microsoft.

Additionally, the treatment group consisted of two subgroups where the order of the test differed to account for order effects. More specifically, while half of the treatment group started in the third round with assigning male names to science concepts, the other half of the treatment group started with linking female names to science concepts. However, Project Implicit reports that the difference in outcomes between these groups, and therefore the order effect, is small.

### *3.1.2. Control group*

The control group started with the Stroop Interference Test (SIT). In this test, subjects have to choose the correct color or word of the text presented. An example of a question of the SIT is shown below in figure 3. The SIT was developed by John Ridley Stroop in 1935 to measure cognitive flexibility and selective attention. Nowadays, the test is used for a broader variety of purposes, such as color blindness. Homack and Riccio (2004) state in their paper that the SIT is most often described as “measuring the subject’s ability to shift between cognitive sets”. In other words, the SIT measures the cognitive ability to restrict a dominant response in favor of using an unusual response.

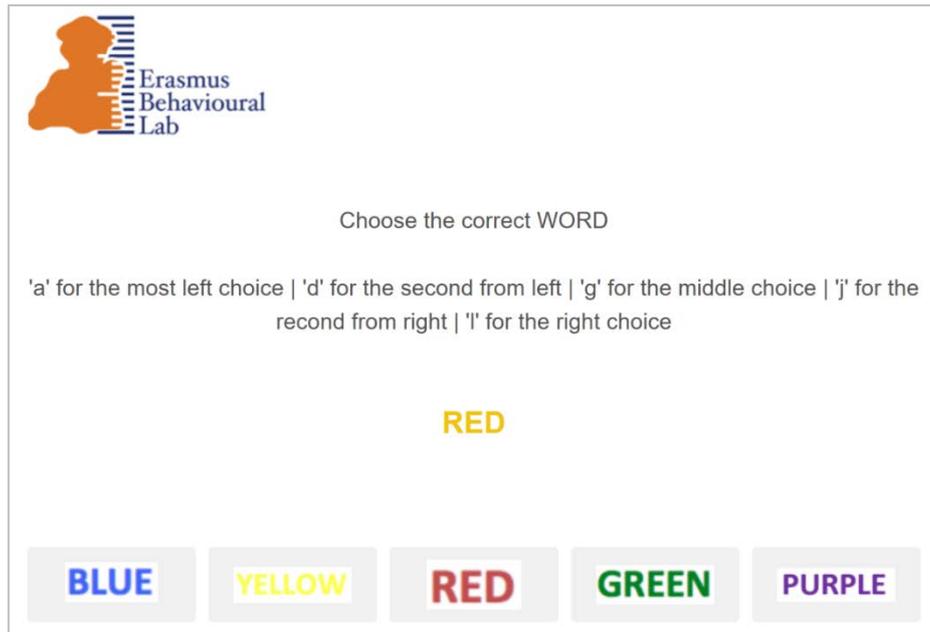


Figure 3. Example of SIT question. The subjects press one of the 5 letters (a, d, g, j, l) to answer the question. Afterwards, they will be automatically be directed to the next similar question. The right answer here is clicking 'g'.

The SIT normally consists of three rounds but was altered for this experiment in such a way, that the subjects spent relatively the same amount of time on the test as the treatment group on the IAT. The SIT was selected for the control group so that the subjects did not suspect that two different tasks were being performed in the room. Both tests were namely performed on a laptop and required only a touch of a key on a regular basis. In addition, the SIT and IAT both require an amount of effort from the brain, making a comparison between the two subject groups more accurately.

In total there are six rounds where the subjects switched between choosing the correct color or word while the difficulty of the task is being increased. The subjects were given their score after submitting all their responses in the end.

### 3.1.3. Perception and behavior statements

The subjects from both the control and treatment groups were asked to assess nine statements shown below. These statements are the dependent variables in this research. In the brackets behind the question is the variable name presented that was used for the analysis. These questions determine the perception and behavior of the subjects on diversity and inclusion (see appendix X for full questionnaire). Questions 1 to 4 and 7 to 9 are statements for which the

subject indicates to what extent he or she agrees with the statement, whereas in questions 5 and 6 the subject chooses one of the three possible answers.

The questions vary between being positively or negatively phrased to decrease the tendency of a subject to agree with a statement when in doubt. This is also referred to as acquiescent response bias. The data is recoded afterwards for a clearer analysis. This will be further elaborated on in section 3.2

- 1) I do not think Diversity & Inclusion (D&I) is an important topic for businesses to grow (DI\_IMP\_BUSINESS)
- 2) I actively promote an inclusive workforce, for example: organizing events / actively starting the discussion with colleagues or partners / actively participating in Girls-day (or similar projects). All focused on D&I. (PROMO)
- 3) Before this session, I was aware of possibly having an unconscious bias in D&I related topics. (SELF\_AWARE\_UnB)
- 4) I do not think that my colleagues are aware of possibly having an unconscious bias in D&I related topics. (COL\_AWARE\_UnB)
- 5) In general, I perceive myself as someone with a relative [*lower/on average/higher*] unconscious bias compared to other people. (SELF\_COMPARED)
- 6) I think my close colleagues have [*a lower/on average/a higher*] unconscious bias compared to me. (COL\_COMPARED)
- 7) I will let others take the Implicit Association Test (IAT)\* this week, because I find becoming aware of your unconscious bias in D&I related topics important. (TAKE\_IAT)
- 8) I do not think it will help the business to make more people aware of their unconscious bias in D&I related topics. (UnB\_BUSINESS)
- 9) *If not in an employee research group:* I will immediately become a member of an ERG after this session.

*If being in an employee research group:* After this session, I will immediately share something interesting, fun, educational or motivation in the ERG Teams group of which I am part of. (ACTION\_ERG)

### 3.1.4. Background questions

Lastly, the subjects answered control questions to account for interpersonal differences and create additional insights for Microsoft NL. This way, Microsoft NL could see where in the organization the resistance towards D&I is the highest, without knowing the specific people that show this behavior. The questions about the amount of years a subject has worked at Microsoft NL (variable *workduration*) and whether the subject is a manager (variable *manager*) are only obtained for Microsoft NL research. These variables will not be further discussed in this paper.

Since the sample size is small, the amount of control questions was restricted. A large amount of control questions with a small sample size could namely lead to a compromised feeling of anonymity of the subjects. Moreover, specifying age groups instead of exact numbers decreases the likelihood that the researcher can trace back who answered what to the questions.

After the background questions and all the subjects had to submit their responses. The control group subjects then received their scores of the SIT. Therefore, having received a positive or negative score did not influence the subjects' assessment to what extent he or she agrees with the perception or behavior statements.

## 3.2. Data

In total 21 people participated and are all employees of Microsoft NL, of which 11 were randomly assigned in the treatment group and 10 in the control group. The data generated through the experiment is presented in table 3 (see next pages) with the variable descriptions. None of the subjects used the option "Refuse to answer". In addition, the data has been re-coded before analyzing. Meaning, the dependent variable statements (statements 1, 4, 5 and 8) that were phrased negatively were re-coded in a way that a more positive response holds a higher value (positively phrased).

Table 3 presents twenty-two variables presented which are all ordinal variables. Moreover, almost all variables have twenty-one observations. Only the variables *Real\_UnB* and *Change\_per* have less observations, namely eleven. The reason for this is because these variables are derived from the questions which were only posed to subjects that took the IAT. However, as stated before, these questions were not used further in this research. Moreover, the

variable *IAT* is a dummy variable with value 1 if the subject was in the treatment group and took the IAT, and with value 0 if the subject was in the control group and took the SIT.

The variables *Imp\_Art*, *Imp\_Science* and *Exp\_UnB* refer to the first three questions asked to all the subjects at the beginning. These variables were however not further used in the analysis in this paper. Future research could use these variables as a randomization check.

*DI\_Imp\_Business*, *Promo*, *Self\_Aware\_UnB*, *Col\_Aware\_UnB*, *Self\_Compare*, *Col\_Compare*, *Take\_IAT*, *UnB\_Business* and *Action\_ERG* are the nine dependent ordinal variables that measure the responses to the nine perception and behavior statements. Since all subjects are asked to assess these statements, the number of observations is 21.

Lastly, the variables *Female*, *Dutch*, *Age\_Group*, *Disability*, *Techrole*, *ERG* and *IAT\_Before* are the control variables. As shown in Table 3, these variables are all dummy variables. Furthermore, the means of these dummies reveal that the majority of the subjects are female, Dutch, do not have a physical disability, do not have a technical role, are part of an ERG or have not taken an IAT before. *Age\_group* refers to which age group the subject belongs and is a ordinal variable with a minimum of 2 and a maximum of 5.

#	Variable		Obs	Mean	Std. Dev.	Min	Max
1	<i>IAT</i>	IAT = 1 if subject was in the treatment group and took the IAT. IAT = 0 if the subject was in the control group and took the SIT.	21	0.52	0.51	0	1
2	<i>Imp_Art</i>	Category of importance of art to the subject	21	2.52	0.98	1	4
3	<i>Imp_Science</i>	Category of importance of science to the subject	21	2.05	0.74	1	4
4	<i>Exp_UnB</i>	Expected unconscious bias of subject	21	3.38	0.86	2	6
5	<i>Real_UnB</i>	Measured unconscious bias of subject through IAT	11	1.19	1.40	0	4
6	<i>Change_per</i>	Change of subject's perception due to score IAT	11	2.29	2.35	0	6
7	<i>DI_Imp_Business</i>	To what extent the subject agrees with the statement: I do not think Diversity & Inclusion (D&I) is an important topic for businesses to grow (1 = not at all, 2 = slightly, 3 = moderately, 4 = very, 5 = extremely)	21	1.19	0.40	1	2
8	<i>Promo</i>	To what extent the subject agrees with the statement: I actively promote an inclusive workforce (1 = not at all, 2 = slightly, 3 = moderately, 4 = very, 5 = extremely)	21	3.95	1.12	2	5
9	<i>Self_Aware_UnB</i>	To what extent the subject agrees with the statement: Before this session, I was aware of possibly having an unconscious bias in D&I related topics (1 = not at all, 2 = slightly, 3 = moderately, 4 = very, 5 = extremely)	21	3.86	1.15	2	5
10	<i>Col_Aware_UnB</i>	To what extent the subject agrees with the statement: I do not think that my colleagues are aware of possibly having an unconscious bias in D&I related topics (1 = not at all, 2 = slightly, 3 = moderately, 4 = very, 5 = extremely)	21	2.90	1.18	1	5
11	<i>Self_Compare</i>	Answer to the question: In general, I perceive myself as someone with a relative [xxx] unconscious bias compared to other people. (1 = lower, 2 = on average, 3 = higher)	21	1.57	0.60	1	3
12	<i>Col_Compare</i>	Answer to the question: In general, I think my close colleagues have [xxx] unconscious bias compared to me (1 = lower, 2 = on average, 3 = higher)	21	2.57	0.51	2	3
13	<i>Take_IAT</i>	To what extent the subject agrees with the statement: I will let others take the Implicit IAT this week, because I find becoming aware of your unconscious bias in D&I related topics important. (1 = not at all, 2 = slightly, 3 = moderately, 4 = very, 5 = extremely)	21	3	1.22	1	5

14	<i>UnB_Business</i>	To what extent the subject agrees with the statement: I do not think it will help the business to make more people aware of their unconscious bias in D&I related topics. (1 = not at all, 2 = slightly, 3 = moderately, 4 = very, 5 = extremely)	21	1.24	0.54	1	3
15	<i>Action_ERG</i>	To what extent the subject agrees with the statement: If not in an ERG: I will immediately become a member of an ERG after this session. If being in an ERG: After this session, I will immediately share something interesting, fun, educational or motivation in the ERG Teams group of which I am part of. (1 = not at all, 2 = slightly, 3 = moderately, 4 = very, 5 = extremely)	21	3.14	1.31	1	5
16	<i>Female</i>	Female = 1 if subject is female, Female = 0 if male	21	0.62	0.50	0	1
17	<i>Dutch</i>	Dutch = 1 if subject is Dutch, Dutch = 0 if otherwise	21	0.71	0.46	0	1
18	<i>Age_Group</i>	To which age group the subject belongs	21	3.14	1.11	2	5
19	<i>Disability</i>	Disability = 1 if the subject has a physical disability, Disability = 0 if not.	21	0.14	0.36	0	1
20	<i>Techrole</i>	Techrole = 1 if the subject holds a technical role within Microsoft NL, Techrole = 0 if not	21	0.43	0.51	0	1
21	<i>ERG</i>	ERG = 1 if the subject is part of an employee research group focused on D&I, ERG = 0 if not.	21	0.81	0.40	0	1
22	<i>IAT_Before</i>	IAT_Before = 1 if the subject has taken an IAT before, IAT_Before = 0 if not	21	0.19	0.40	0	1

Table 3. Summary statistics on all variables gathered from the subjects on both days.

### 3.3. Hypotheses & Models

Aside from the main objective of this paper to create a methodology which measures the effectiveness of the IAT on creating awareness, this sub-chapter focusses on how the data can be analyzed once this study has been conducted in a larger setting.

#### 3.3.1. Overall analysis

Before the data is analyzed in more detail, nine graphs are generated to show the cumulative probabilities of the responses per dependent variables for both the control group and treatment group. Moreover, a MANOVA test is conducted to see whether there is a statistically significant difference in any of the dependent variables between the treatment and control group.

#### 3.3.2. Single questions analyses

Secondly, the nine questions are evaluated separately to determine whether taking the IAT has an effect on the extent to which the subject agrees with the single statements. For this, the following model 1 is used for each of the nine perception and behavior statements:

$$\mathbf{Statement}_{ij} = \beta_0 + \beta_1 IAT_i + \beta_2 Female_i + \beta_3 Dutch_i + \beta_4 Age\_Group_i + \beta_5 Disability_i + \beta_6 Techrole_i + \beta_7 ERG_i + \beta_8 IAT\_Before_i + \varepsilon_i \quad (1)$$

Where *Statement* is the dependent variable of which ‘i’ stands for the subject and ‘j’ refers to one of the nine statements (1 = *DI\_Imp\_Business*, 2 = *Promo*, 3 = *Self\_Aware\_UnB*, 4 = *Col\_Aware\_UnB*, 5 = *Self\_Compare*, 6 = *Col\_Compare*, 7 = *Take\_IAT*, 8 = *UnB\_Business* and 9 = *Action\_ERG*). The variable of interest is *IAT*, which is a dummy variable per individual with the value of 1 when the subject has taken the Implicit Association Test, and the value of 0 if not. Moreover, the remaining variables *Female*, *Dutch*, *Age\_Group*, *Disability*, *Techrole*, *ERG* and *IAT\_Before* are all ordinal control variables (see table 3 for detailed explanation of the variables). The  $\varepsilon_i$  is the residual per subject ‘i’. The hypothesis that is tested per question is the following:

$$H0: \beta_1 = 0$$

$$H1: \beta_1 \neq 0$$

Meaning, the null hypothesis will be rejected when taking the IAT has a significant effect on the answered category of the question<sup>5</sup>. This model will be tested in STATA through an ordered logit regression since both the dependent and independent variables are ordered.

Additionally, a Mann Whitney-U test is performed in STAT on the nine statements to compare the ranks of both the treatment group as the SIT control group.

### 3.3.2.1. Principal factor analysis

Thirdly, the perception and behavior statements are analyzed through the principal factor analysis to test for latent variables. Latent variables are variables that cannot directly be measured through one variable. They can be observed through the interrelationships between a set of dependent variables. More specifically, the latent variable affects the subject to assess the statements (on which the latent variable has an effect) in a certain way. An example of a latent variable that could be observed through the nine statements in this data is: negative perception of others on D&I. In addition, the principal factor analysis is different from the principal component analysis (PCA). The PCA serves merely to reduce the data to the optimal amount of components, the optimal number of measured variables for each component and the optimal weights.

Finally, the latent variables that are found in the data are interpreted and tested through a standard OLS regression. This way, we can observe whether certain behaviors or other patterns are statistically different after taking an IAT. The regression model used here is similar to the previous model (1):

$$\mathbf{LatentVariable}_{ji} = \gamma_0 + \gamma_1 \mathbf{IAT}_i + \gamma_2 \mathbf{Female}_i + \gamma_3 \mathbf{Dutch}_i + \gamma_4 \mathbf{Age\_Group}_i + \gamma_5 \mathbf{Disability}_i + \gamma_6 \mathbf{Techrole}_i + \gamma_7 \mathbf{ERG}_i + \gamma_8 \mathbf{IAT\_Before}_i + \varepsilon_i \quad (2)$$

Where *LatentVariable* is the latent variable found in the data of which ‘j’ refers to the number of the latent variables found and ‘i’ for the subject. The hypothesis tested for this model is the following:

$$\mathbf{H0: } \gamma_1 = \mathbf{0}$$

$$\mathbf{H1: } \gamma_1 \neq \mathbf{0}$$

---

<sup>5</sup> p-value of IAT variable < 0.05

More specifically, the null hypothesis will be rejected when taking the IAT has a significant effect on the latent variable.

## 4. Results

### 4.1. Overall analysis

As can be seen in the graphs below (figures 4 – 12), the cumulative probabilities of the assessments per dependent variable do not show much difference between the group that took the IAT before assessing the statements compared to the group that took the SIT before assessing the statements. The small difference is shown as the red (treatment group) and blue (control group) lines lie relatively close to each other. Only the variables *Action\_ERG*, *PROMO* and *Self\_Aware\_UNB* indicate some difference in the probability of the outcomes. The cumulative probabilities for the treatment group in *Action\_ERG* are scaled to the right, meaning that the chance that these subjects reported a higher value is greater compared to the control group subjects. The variable *PROMO* shows the same progress. However, the variable *Self\_Aware\_UNB* shows the opposite.

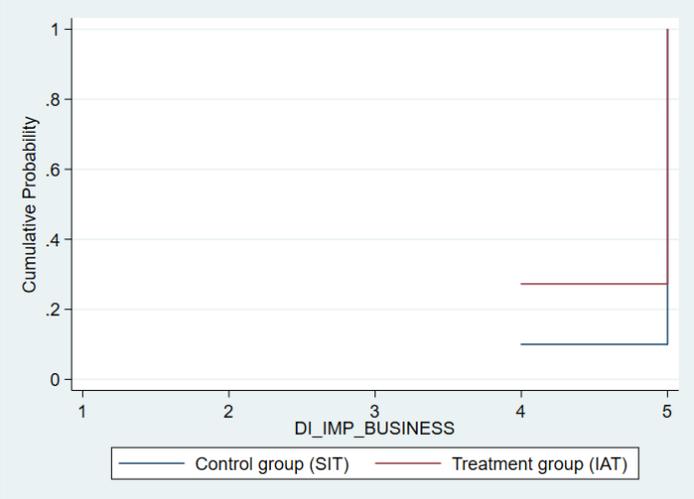


Figure 4. Cumulative probability of *DI\_IMP\_BUSINESS* for treatment and control group

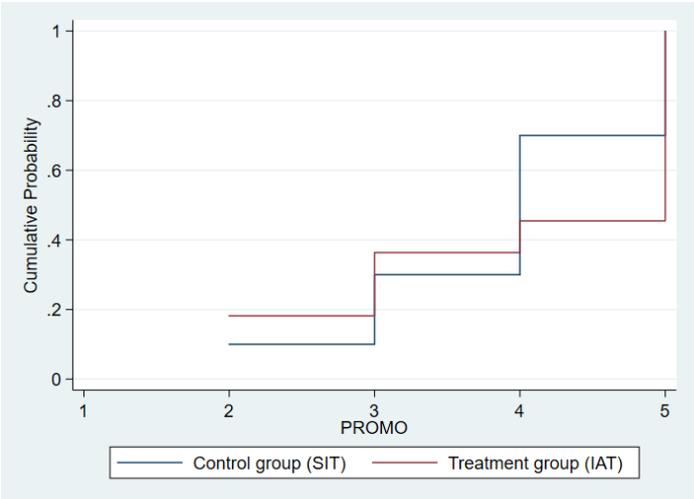


Figure 5. Cumulative probability of *PROMO* for treatment and control group

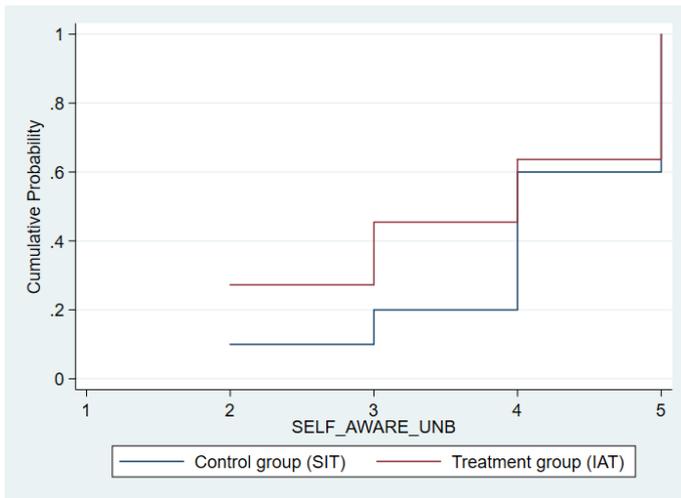


Figure 6. Cumulative probability of SELF\_AWARE for treatment and control group

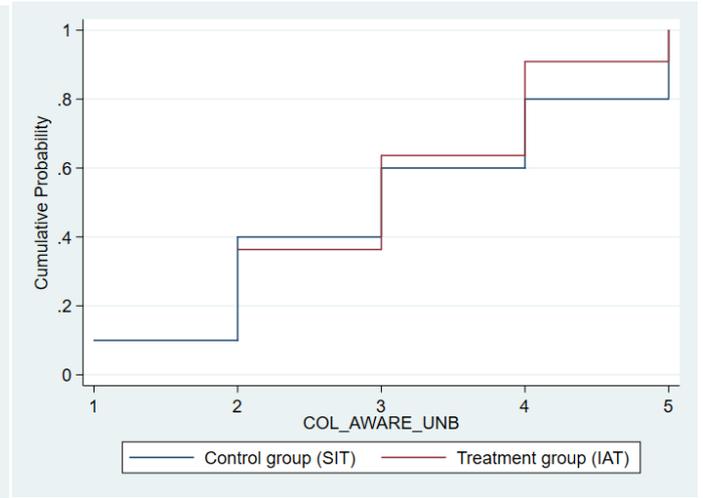


Figure 7. Cumulative probability of COL\_AWARE for treatment and control group

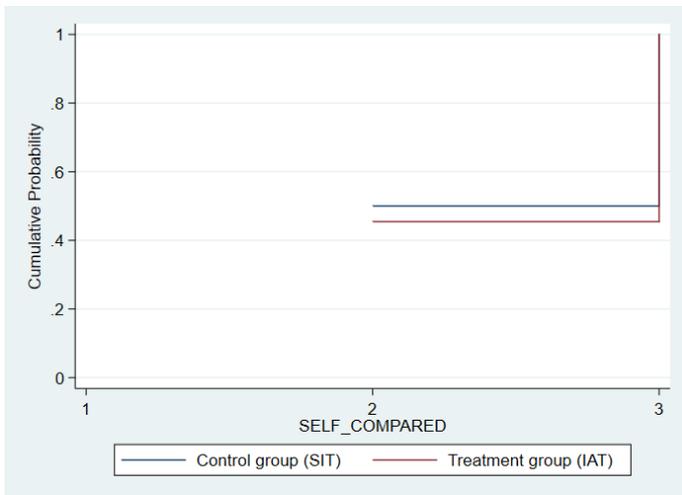


Figure 8. Cumulative probability of SELF\_COMPARED for treatment and control group

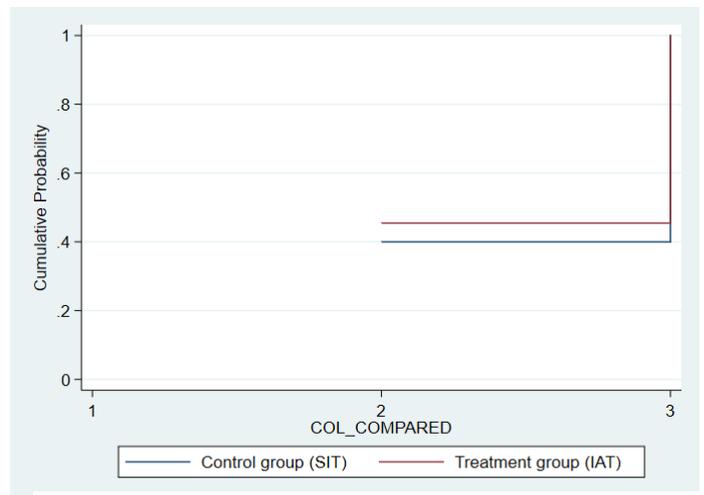


Figure 9. Cumulative probability of COL\_COMPARED for treatment and control group

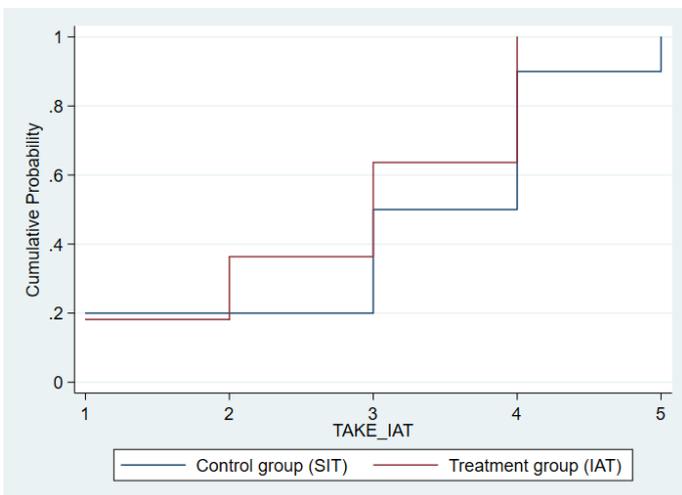


Figure 10. Cumulative probability of TAKE\_IAT for treatment and control group

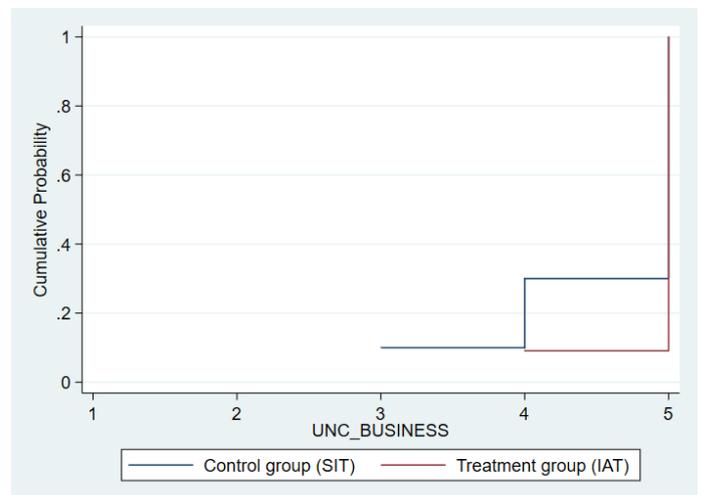


Figure 11. Cumulative probability of UNC\_BUSINESS for treatment and control group

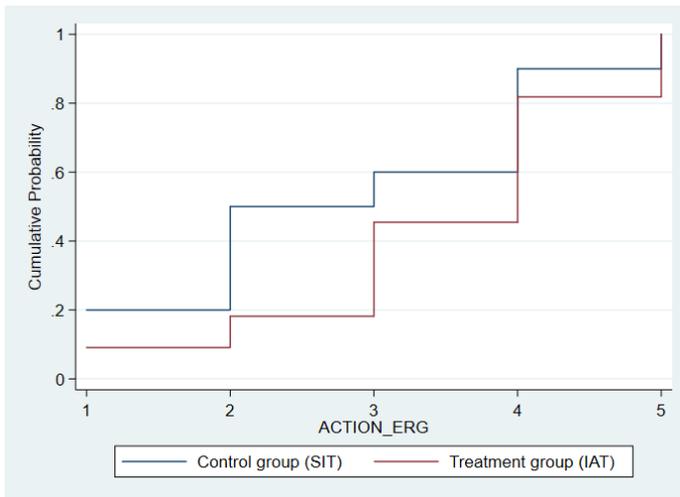


Figure 12. Cumulative probability of ACTION\_ERG for treatment and control group

Furthermore, the MANOVA test shows results in line with these graphs. Namely, figure D (appendix) indicates p-values of 0.1387. In other words, there is no significant difference between the treatment and control group found in any of the dependent variables.

#### 4.2. Single question analyses

The results from the overall analysis don't show any significant effect. However, this section aims to provide an analysis framework for when this study is conducted in a larger setting. Here, the effect of the IAT is tested on the nine behavioral questions individually with model (1). In Table 4 (page 28) the results are shown for using an ordered logit model. A value of 1.692 for the variable IAT is the ordered log-odds estimate for when the subject took the IAT, compared to taking the SIT, on the expected DI\_IMP\_BUSINESS level given the other variables are held constant in the model. Being in the treatment group and therefore taking the IAT would result in a 0.0532 unit increase in the ordered log-odds of being in a higher DI\_IMP\_BUSINESS category while the other variables in the model are held constant.

Furthermore, table 4 (page 28) shows that the null hypothesis ( $H_0: \beta_1 = 0$ ) holds for all questions since none of the parameters for IAT is statistically significant. In other words, there is no evidence that supports the hypothesis that taking the IAT impacts the extent to which a subject agrees with the perception and behavior statements.

In addition, Table 5 on the next page shows that also here no significant difference between the two groups was found through the Mann Whitney U test, because no p-value is smaller than 0.05.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Variable	DI_IMP_ BUSINESS	PROMO	SELF_AWARE_ UNB	COL_AWARE_ UNB	SELF_COMPAR ED	COL_COMPAR ED	TAKE_IAT	UNC_BUSINESS	ACTION_ ERG
z-value	0.98	-0.45	0.77	0.000	-0.20	0.25	0.77	-1.24	-1.07
p-value	0.33	0.66	0.44	1.0000	0.83	0.81	0.44	0.22	0.28

Table 5. Mann Whitney U test results of all nine questions between the treatment (IAT) and control (SIT) group. None of the p-values is smaller than 0.05, meaning that no statistical significant difference is found between the ranks of the control and treatment group.

VARIABLES	(1) Question 1 DI_IMP_ BUSINESS	(2) Question 2 PROMO	(3) Question 3 SELF_ AWARE_UNB	(4) Question 4 COL_ AWARE_UNB	(5) Question 5 SELF_ COMPARED	(6) Question 6 COL_ COMPARED	(7) Question 7 TAKE_ IAT	(8) Question 8 UNC_ BUSINESS	(9) Question 9 ACTION_ERG
<i>IAT</i>	-1.69 (1.87)	0.46 (0.93)	-0.46 (0.88)	0.61 (0.90)	0.004 (1.20)	-0.33 (1.12)	-1.14 (1.00)	582.3 (6.903e+08)	0.94 (0.89)
<i>Female</i>	2.28 (1.88)	-0.79 (1.47)	0.81 (1.31)	3.77** (1.64)	0.47 (1.73)	16.71 (2,93)	0.06 (1.52)	707.6 (9.204e+08)	-0.34 (1.44)
<i>Dutch</i>	-0.60 (2.54)	-1.93 (1.34)	1.36 (1.23)	4.75*** (1.54)	1.24 (1.65)	-1.52 (1.59)	-1.96 (1.26)	-226.6 (2.301e+08)	-0.96 (1.11)
<i>Age_Group</i>	0.002 (1.17)	0.28 (0.42)	0.33 (0.45)	0.24 (0.45)	-0.045 (0.65)	0.62 (0.73)	1.14** (0.56)	177.7 (2.301e+08)	0.81* (0.46)
<i>Disability</i>	19.95 (8,76)	-1.06 (1.82)	2.59* (1.57)	1.73 (1.63)	0.61 (1.92)	18.01 (2,94)	0.10 (1.74)	124.8 (2.301e+08)	0.89 (1.56)
<i>Techrole</i>	-1.08 (1.70)	0.33 (0.91)	2.25* (1.15)	2.45** (1.09)	-0.43 (1.33)	1.47 (1.34)	1.60 (1.06)	229.9 (2.301e+08)	0.37 (0.87)
<i>ERG</i>	0.79 (2.38)	4.16** (1.87)	-4.34** (1.94)	-3.53** (1.71)	-19.5 (6,08)	-18.18 (2,94)	1.00 (1.80)	-935.8 (1.150e+09)	0.58 (1.54)
<i>IAT_before</i>	17.24 (7,88)	0.48 (1.20)	1.35 (1.19)	-0.80 (1.29)	-19.36 (6,80)	-0.86 (1.65)	1.51 (1.52)	-355.1 (4.602e+08)	-0.43 (1.24)
<i>/cut1</i>	-1.13 (4.14)	0.30 (2.06)	-1.94 (2.01)	0.45 (1.91)	-18.59 (6,08)	-0.48 (2.35)	1.00 (2.13)	-538.5 (4.637e+08)	0.50 (1.87)
<i>/cut2</i>		1.84 (2.14)	-0.85 (2.00)	4.16** (2.02)			1.71 (2.13)	104.3 (2.301e+08)	1.95 (1.96)
<i>/cut3</i>		3.22 (2.24)	0.97 (1.95)	5.60*** (2.15)			3.62* (2.19)		2.99 (2.03)
<i>/cut4</i>				7.75*** (2.49)			8.67*** (3.28)		4.96** (2.15)
Observations	21	21	21	21	21	21	21	21	21

Table 4. Ordered logit regression results of all dependent variables. The variable of interest, IAT, does not show any significance for any of the nine statements

### 4.3. Principal factor analysis

Finally, the principal factor analysis executed in the figures E (appendix) and the screeplot below (figure 13) show that three latent variables are found, using the rule of thumb eigenvalues  $> 1$ .

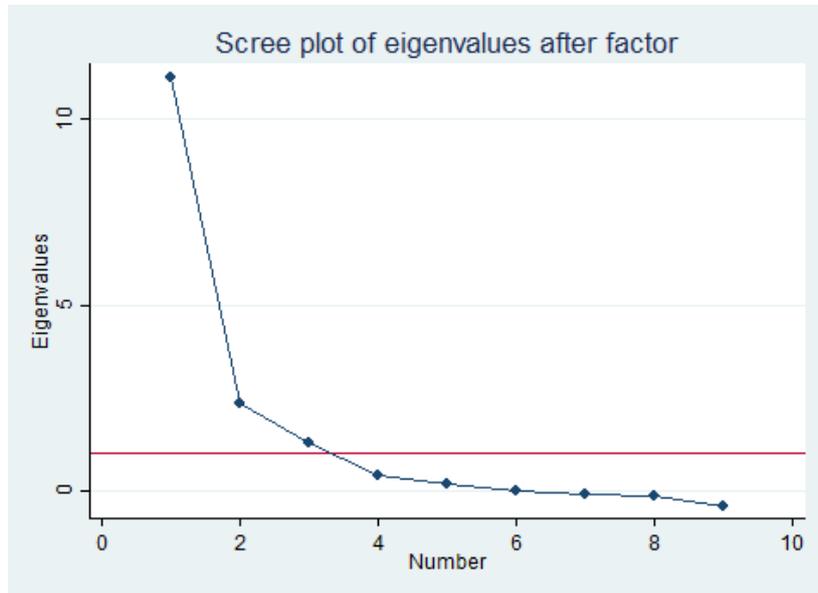


Figure 13. Graph showing the eigenvalues after iterated principal factor analysis.

In table 6 below, the factor loadings of these latent variables are presented. The first latent variable is interpreted as ‘comparison to colleagues’ as the highest variance for this latent variable comes from the variable *Col\_Compare*. Furthermore, the second latent variable is identified as ‘behavior’, since the behavior focused variables *Promo*, *Take\_IAT* and *Action\_ERG* contribute the most to the variance of the second factor. Lastly, the third latent variable is labeled as ‘awareness unconscious bias’ as the variance of this direction is the highest for the variables *Col\_Aware*, *Self\_Aware*, *DI\_Imp\_Business* and *Action\_ERG*.

Variable	Factor 1	Factor 2	Factor 3	Uniqueness
<i>DI_IMP_BUSINESS</i>		0.49	0.38	0.61
<i>PROMO</i>		0.74		0.42
<i>SELF_AWARE_UNB</i>		-0.44	0.47	0.59
<i>COL_AWARE_UNB</i>			0.87	0.15
<i>SELF_COMPARED</i>				0.98
<i>COL_COMPARED</i>	3.33			-10.06
<i>TAKE_IAT</i>		0.79		0.35
<i>UNC_BUSINESS</i>			0.3804	0.80
<i>ACTION_ERG</i>		0.78		0.37
Blanks represent abs(loadings)<0.3				

Table 6. Factor loadings of principal factor analysis

Additionally, these three latent variables are analyzed through a standard OLS regression on the independent variable IAT and control variables (see model 2). The output is presented in table 7, showing that for all three latent variables the null hypothesis holds. In other words, no evidence is found that taking the IAT before assessing the diversity and inclusion statements has a significant effect on someone's self-comparison with colleagues, behavior or awareness of the unconscious bias.

VARIABLES	LV 1 Compared to colleagues	LV 2 Behavior	LV 3 Awareness unconscious bias
IAT	-0.29 (2.31)	0.18 (0.51)	-0.16 (0.36)
Female	9.49** (3.70)	-1.27 (0.81)	1.76*** (0.57)
Dutch	2.55 (2.91)	-1.06 (0.64)	1.41*** (0.45)
Age_Group	-0.02 (1.16)	0.23 (0.25)	0.07 (0.18)
Disability	8.48* (4.46)	-1.10 (0.98)	1.06 (0.69)
Techrole	2.94 (2.37)	-0.38 (0.52)	0.87** (0.37)
ERG	-5.20 (4.09)	1.54 (0.90)	-1.02 (0.63)
IAT_BEFORE	-4.08 (3.21)	0.63 (0.71)	-0.39 (0.50)
Constant	-4.96 (4.86)	-0.32 (1.07)	-1.84** (0.75)
Observations	21	21	21
R-squared	0.42	0.40	0.62

Standard errors in parentheses

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table 7. Results of the latent variable analyses. None of the parameters of the variable of interest (IAT) is statistically significant.

## 5. Discussion, limitations and future research

### 5.1. Discussion

The aim of this research is to create a quantitative design for companies to assess whether the IAT can be used to create awareness for D&I within their organization. The research question central in this study is: Does completing an IAT on discrimination have an effect on a person's awareness and behavior towards D&I? The developed design was tested as a pilot at Microsoft NL. The data generated from this methodology is then analyzed and proposed to be analyzed for future research in three ways: overall analysis, single question analyses and principal factor analysis. These three types of analysis all gave the same result in this pilot experiment, namely no evidence to conclude that completing an IAT on discrimination has an effect on a person's awareness and behavior towards D&I.

Regardless of the not significant results, the contribution of this paper is the developed design. With this paper, organizations can replicate or increase the scope of this study without much required resources. Moreover, this design includes randomization and a control group to resolve self-selection. Additionally, this design generates quantitative data that can then be analyzed through three proposed types of analysis. This way, organizations can assess the effectiveness of an IAT focused on discrimination, on the D&I awareness of their employees. In addition, organizations create an update of the level of unconscious bias<sup>6</sup> within their companies. This can help these organizations to stress the importance of a more diverse and inclusive workforce to their employees and further develop their D&I strategy.

### 5.2. Limitations and additional considerations

This study includes several caveats that are expected to have influenced the results. The first limitation of this study is the sample size. The number of subjects that participated in this experiment was 21. However, using the Cochran formula<sup>7</sup> with a confidence level of 95% and a population of 500<sup>8</sup>, a sample size of 225 is suggested to be a better fit. This much lower sample size decreases the statistical power, increasing the likelihood of not finding the true effects of the population (Ellis, 2010).

---

<sup>6</sup> A KPI for inclusion

<sup>7</sup>  $n = n_0 / (1 + (n_0 - 1/N))$  where  $n_0 = (Z^2 * pq / e^2)$

<sup>8</sup> Employees of Microsoft NL

Moreover, of the 21 subjects 81% was part of an employee research group (ERG). The ERG's in Microsoft NL are focused on increasing D&I within the company. This means that these subjects are more likely aware of their unconscious bias compared to their colleagues who are not part of an ERG. Additionally, these subjects share characteristics that could have possibly led to a smaller effect in outcome. For future research, it is therefore recommended to include a larger and more diverse sample.

Another limitation of this study is that the whole experiment was not programmed within the Qualtrics environment. The treatment group took the IAT on the project implicit website, and was asked to come back to the Qualtrics experiment environment to report their score afterwards. Not only does this create the opportunity for the subjects to lie about their scores, it also potentially increases confusion about the process in the subjects. Therefore, this study has not included the IAT scores in the tested models, to account for the possibility that the subjects lied about their scores. However, all the subjects finished the test and reported their scores showing that the redirection had no significant effect on the understanding of the process.

Furthermore, a switch between positive and negative wording in the behavior and perception statements was chosen to account for acquiescent response bias. However, Schriesheim and Hill (1981) argue that there is no empirical evidence that supports this statement. More specifically, in their study they found that it is not advisable to shift between positive and negative wording, as this decreases the accuracy of responses.

Besides these findings, I found that conducting this experiment at Microsoft NL was easy to implement. The only difficulty that I experienced while conducting this experiment is to get all the subjects for half an hour in the same, big enough, room. Future research is therefore suggested to plan this experiment at least 2 months in advance, to ensure a larger sample size and find a suitable location.

## 6. Conclusion

Organizations worldwide increasingly use the IAT to develop awareness on D&I. However, the effectiveness of the IAT on creating this awareness is inconclusive. This study provides a quantitative design to fill this gap in literature and to measure the effectiveness of an IAT<sup>9</sup> within a company. Furthermore, this design was tested through a pilot at Microsoft NL. This

---

<sup>9</sup> Focused on discrimination

pilot did not support the hypothesis that the taking an IAT will create awareness on D&I. However, recommended is to increase the scope of this experiment to measure the true effects of the population.

## 7. References

- Atewologun, D., Cornish, T., & Tresh, F. (2018). Unconscious bias training: An assessment of the evidence for effectiveness. *Equality and Human Rights Commission Research Reports Series*.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74(5), 1252-1265.
- Bertrand, M., Black, S. E., Jensen, S., & Lleras-Muney, A. (2014). Breaking the glass ceiling? The effect of board quotas on female labor market outcomes in Norway (No. w20256). *National Bureau of Economic Research*.
- Boston Consulting Group (2017). Retrieved on 02-04-2019 from: <https://www.bcg.com/en-us/publications/2017/diversity-at-work.aspx>
- Busenitz, L., & Barney, J. (1997). Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *Journal of Business Venturing*, 12(1), 9-30
- Cialdini, R. B., & Sagarin, B. J. (2005). Principles of Interpersonal Influence. Ferris State University. Retrieved on 07-06-2019 from <https://www.ferris.edu/htmls/administration/president/diversityoffice/definitions.htm>
- Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human perception and performance*, 3(4), 552.
- Frantz, C. M., Cuddy, A. J., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin*, 30(12), 1611-1624.
- Furst, S. A., & Cable, D. M. (2008). Employee resistance to organizational change: Managerial influence tactics and leader-member exchange. *Journal of Applied psychology*, 93(2), 453.
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of general internal medicine*, 22(9), 1231-1238.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Journal of Personality and Social Psychology*, 102, 4-27.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (1998). Retrieved on 23-08-2019 from: <https://www.projectimplicit.net/about.html>

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6), 1464.

Homack, S., & Riccio, C. A. (2004). A meta-analysis of the sensitivity and specificity of the Stroop Color and Word Test with children. *Archives of clinical Neuropsychology*, 19(6), 725-743.

Hunt, V., Yee, L., & Prince, S. (2018). Retrieved on 22-08-2019 from: <https://www.mckinsey.com/business-functions/organization/our-insights/delivering-through-diversity>

Jajodia, A., & Earleywine, M. (2003). Measuring alcohol expectancies with the implicit association test. *Psychology of Addictive Behaviors*, 17(2), 126.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the implicit association test. *Journal of personality and social psychology*, 81(5), 774.

Lee, A. J. (2005). Unconscious bias theory in employment discrimination litigation. *Harv. CR-CLL Rev.*, 40, 481.

Maison, D., Greenwald, A. G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes.

Microsoft. (26-01-2019). Retrieved on 22-07-2019 from: <https://www.microsoft.com/en-us/diversity/inside-microsoft/default.aspx#epgDivFocusArea>

Miller Jr, D. P., Spangler, J. G., Vitolins, M. Z., Davis, M. S. W., Ip, E. H., Marion, G. S., & Crandall, S. J. (2013). Are medical students aware of their anti-obesity bias?. *Academic medicine: journal of the Association of American Medical Colleges*, 88(7), 978.

Morris, K. A., & Ashburn-Nardo, L. (2009). The Implicit Association Test as a class assignment: Student affective and attitudinal reactions. *Teaching of Psychology*, 37(1), 63-68.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: personalizing the IAT. *Journal of personality and social psychology*, 86(5), 653.

Pouwels, B., & Henderikse, W. (2017). *Bedrijvenmonitor Topvrouwen 2017*.

Rijksoverheid. (2017). Retrieved on 22-08-2019 from: <https://www.rijksoverheid.nl/actueel/nieuws/2017/03/24/wettelijk-streefcijfer-mannen-en-vrouwen-in-bestuur-ondernemingen-van-kracht>

Samson, A. (2018). *The Behavioral economics guide*. London: Behavioral Science Solutions, Ltd.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, 1(1), 7-59.

Schriesheim, C. A., & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and psychological measurement*, 41(4), 1101-1114

Sheeran, P., & Webb, T. L. (2016). The intention–behavior gap. *Social and personality psychology compass*, 10(9), 503-518.

Starbucks Canada (4-5-2018). Retrieved on 22-08-2019 from: [https://twitter.com/StarbucksCanada/status/992500325687873536/photo/1?ref\\_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed&ref\\_url=http%3A%2F%2Ftheconversation.com%2Fstarbucks-and-the-impact-of-implicit-bias-training-96491](https://twitter.com/StarbucksCanada/status/992500325687873536/photo/1?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed&ref_url=http%3A%2F%2Ftheconversation.com%2Fstarbucks-and-the-impact-of-implicit-bias-training-96491)

Tipper, J. (2004). How to increase diversity through your recruitment practices. *Industrial and commercial training*, 36(4), 158-161.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, England: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin*, 133(5), 859.

Vanderbilt.edu. (2-5-2019). Retrieved on 5-6-2019 from <https://www.vanderbilt.edu/diversity/unconscious-bias/>

Vriesinga (17-5-2019). Retrieved on 6-6-2019 from <https://www.nrc.nl/nieuws/2019/05/17/je-werkt-beter-wanneer-je-licht-oncomfortabel-bent-a3960704>

## 8. Appendix

Figure A: IAT example 1

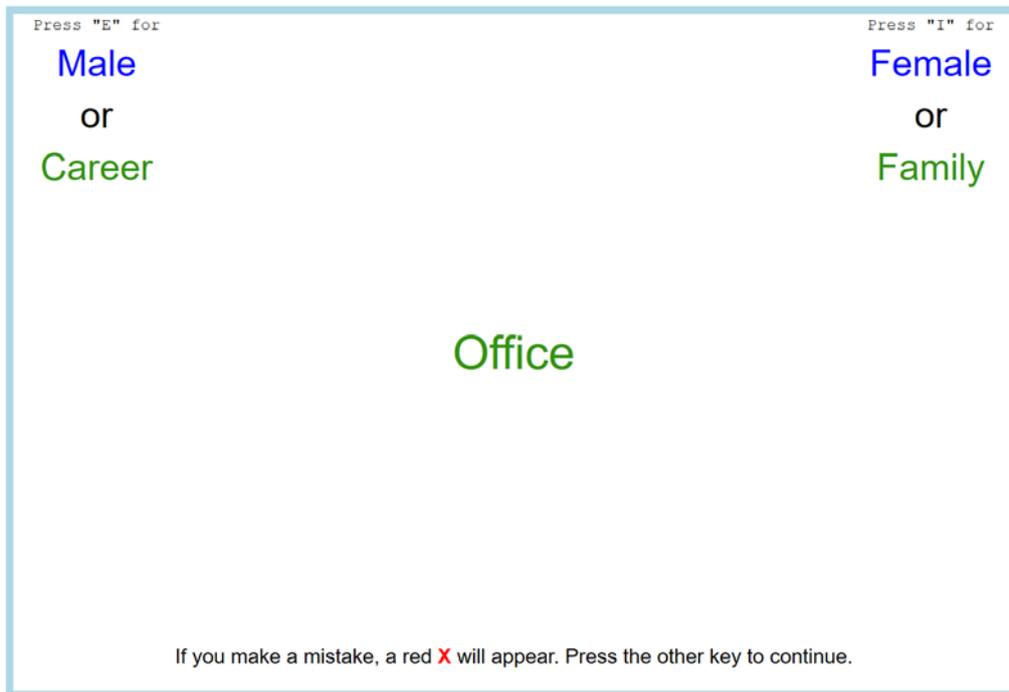


Figure 3. Possible round three of the Implicit Association Test

Figure B: IAT example 2

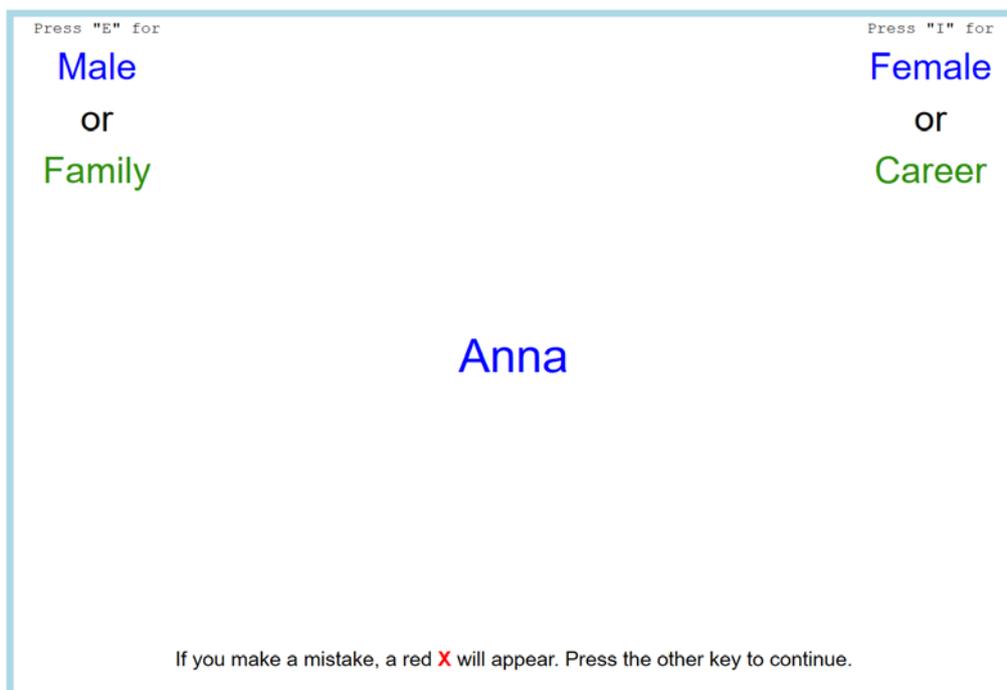


Figure 4. Round five of the Implicit Association Test

Figure C. Treatment group instructions



You are now going to make the **Gender-Science Implicit Association Test (IAT)**. For this, you will be redirected to another website.

Please follow the following steps carefully:

1. Open another tab.
2. Copy the following link in this new tab: <https://implicit.harvard.edu/implicit/takeatest.html>
3. Click on the blue text **'I wish to proceed'**
4. On the left, you see all the possible IAT's. Please choose the box **Gender-Science IAT**. Be careful **not** to choose Gender-Career IAT.
5. Click on the 'Continue' button.
6. Before (or after) the test starts, you might be asked (not always) to answer some questions. You can choose to click 'Decline to Answer' in the bottom right to speed up the process. This will not affect my research.
7. Make the test.
8. Remember your score. Keep this tab open in case you forget your score.
9. Come back to this survey.
10. Click the next button to submit your score in the next question.



Figure3. Screen displayed to treatment group before starting the IAT.

Figure D: MANOVA output

Number of obs = 21

W = Wilks' lambda      L = Lawley-Hotelling trace  
P = Pillai's trace      R = Roy's largest root

Source	Statistic	df	F(df1, df2) =	F	Prob>F	
IAT	W	0.3794	1	9.0	11.0	2.00 0.1387 e
	P	0.6206		9.0	11.0	2.00 0.1387 e
	L	1.6357		9.0	11.0	2.00 0.1387 e
	R	1.6357		9.0	11.0	2.00 0.1387 e
Residual		19				
Total		20				

e = exact, a = approximate, u = upper bound on F

Figure E: Principal factor analysis

Factor analysis/correlation      Number of obs = 21  
Method: iterated principal factors      Retained factors = 8  
Rotation: (unrotated)      Number of params = 36

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	2.78883	1.20596	0.3922	0.3922
Factor2	1.58287	0.47711	0.2226	0.6149
Factor3	1.10576	0.22760	0.1555	0.7704
Factor4	0.87816	0.50455	0.1235	0.8939
Factor5	0.37361	0.05129	0.0525	0.9464
Factor6	0.32232	0.27185	0.0453	0.9918
Factor7	0.05047	0.04203	0.0071	0.9989
Factor8	0.00844	0.00871	0.0012	1.0000
Factor9	-0.00027	.	-0.0000	1.0000

LR test: independent vs. saturated:  $\chi^2(36) = 65.74$  Prob> $\chi^2 = 0.0018$

## Appendix X: Full Qualtrics experiment

# Experiment MSc Thesis

---

### Start of Block: Introduction

Introduction

Hello!

Thank you for participating in this experiment. During this session I kindly ask you to silence your phone and to update your Teams status to 'do not disturb'.

Please remember:

- 1) The purpose of this experiment is for scientific research only
- 2) All the data will be anonymized

These two considerations are there to remind you that you can and should only answer what you really think. Do not answer the questions with what you think people will want to hear.

Please read all the questions and answers carefully.

You can now click arrow button to start.

### End of Block: Introduction

---

### Start of Block: Controls before

Q1.1 How important is **art** to you?

- Extremely important (1)
  - Very important (2)
  - Moderately important (3)
  - Slightly important (4)
  - Not at all important (5)
-

Q1.2 How important is **science** to you?

- Extremely important (1)
  - Very important (2)
  - Moderately important (3)
  - Slightly important (4)
  - Not at all important (5)
- 

Q1.3

**Research has shown that people often have an unconscious bias relating to Diversity & Inclusion (D&I) topics. Unconscious biases are learned stereotypes that are automatic, unintentional, deeply engrained, universal and able to influence behavior. Moreover, people can have an unconscious bias while being aware of it.**

Do you think you have an unconscious bias relating to gender and science or art? I think I have a...

- strong automatic association of Males with Science and Females with Liberal Arts (1)
- moderate automatic association of Males with Science and Females with Liberal Arts (2)
- slight automatic association of Males with Science and Females with Liberal Arts (3)
- no automatic association (4)
- slight automatic association of Females with Science and Males with Liberal Arts (5)
- moderate automatic association of Females with Science and Males with Liberal Arts (6)
- strong automatic association of Females with Science and Males with Liberal Arts (7)

End of Block: Controls before

---

Start of Block: IAT

## TREATMENT GROUP<sup>10</sup>

You are now going to make the **Gender-Science Implicit Association Test (IAT)**. For this, you will be redirected to another website.

Please follow the following steps carefully:

1. Open another tab.
2. Copy the following link in this new tab: <https://implicit.harvard.edu/implicit/takeatest.html>
3. Click on the blue text '**I wish to proceed**'
4. On the left, you see all the possible IAT's. Please choose the box **Gender-Science IAT**. Be careful **not** to choose Gender-Career IAT.
5. Click on the 'Continue' button.
6. Before (or after) the test starts, you might be asked (not always) to answer some questions. You can choose to click 'Decline to Answer' in the bottom right to speed up the process. This will not affect my research.
7. Make the test.
8. Remember your score. Keep this tab open in case you forget your score.
9. Come back to this survey.
10. Click the next button to submit your score in the next question.

End of Block: IAT

---

Start of Block: Color test

## CONTROL GROUP<sup>11</sup>

You will now start the Stroop Effect Test. This test was developed to measure someone's cognitive ability in terms of flexibility. Also, some argue that possible brain damage (and color blindness) can be detected through this test.

In the first part you will be asked to choose the correct WORD presented. So you do not have to do anything with the color of the word.

In the second part, you will be asked to choose the correct COLOR of the word stated. This switch will be indicated.

---

<sup>10</sup> This was not shown to the subject

<sup>11</sup> This was not shown to the subject

Keep your fingers on:  
press 'a' for the left choice  
press 'h' for the middle choice  
press 'l' for the right choice

You will receive your score by the end of this test.

You can now click the next button to start.

Good luck!

---

Page Break

All the rounds

---

Page Break

You are now done with the Stroop Effect Test, your score will become visible after finishing this experiment.

As you may have experienced yourself, it takes for most people longer to state the colors of the words, rather than to read the text there are printed in. This is interesting, because the incongruence is essentially the same across both lists (in round 2, 4, 5 and 6 the words are shown in the wrong color). It appears we are more influenced by the physical text than the text color.

But why does this happen? What this reveals is that the brain can not help but read. As habitual readers, we encounter and comprehend words on such a persistent basis that the reading occurs almost effortlessly, whereas declaration of a color requires more cognitive effort. When there is a conflict between these two sources of information, our cognitive load is increased, and our brains have to work harder to resolve the required difference. Performing these tasks (preventing reading, processing word color, and resolving information conflict) ultimately slows down our responses, and makes the task take longer.

---

Page Break

QA1 What was your score of the IAT?

- Strong automatic association of Males with Science and Females with Liberal Arts (1)
- Moderate automatic association of Males with Science and Females with Liberal Arts (2)
- Slight automatic association of Males with Science and Females with Liberal Arts (3)
- Little to no automatic association (4)
- Slight automatic association of Females with Science and Males with Liberal Arts (5)
- Moderate automatic association of Females with Science and Males with Liberal Arts (6)
- Strong automatic association of Females with Science and Males with Liberal Arts (7)
- Refuse to answer (8)

*Display This Question: in the Treatment group and took the IAT*

QA2

To what extend did the IAT-score you received change (positive or negative) your view of yourself towards Diversity & Inclusion? It changed:

- Extremely positive (1)
- Moderately positive (2)
- Slightly positive (3)
- Neither positive nor negative (4)
- Slightly negative (5)
- Moderately negative (6)
- Extremely negative (7)
- Refuse to answer (8)

End of Block: Q only if taken IAT

---

Start of Block: Questions

PART 1

Q2.1 – Q2.4

To what extent do you agree with the following statements?

	Not at all (1)	Slightly (2)	Moderately (3)	Very (4)	Extremely (5)	Refuse to Answer (6)
<b>1) I do not think Diversity &amp; Inclusion (D&amp;I) is an important topic for businesses to grow (1)</b>	<input type="radio"/>					
<b>2) I actively promote an inclusive workforce, for example such as: organizing events / actively starting the discussion with colleagues or partners / actively participating in girls day (or similar projects). All focused on D&amp;I (2)</b>	<input type="radio"/>					
<b>3) Before this session, I was aware of possibly having an unconscious bias in D&amp;I related topics (3)</b>	<input type="radio"/>					
<b>4) I do not think that my colleagues are aware of possibly having an unconscious bias in D&amp;I related topics (4)</b>	<input type="radio"/>					

Page Break

Q2.5

In general, I perceive myself as someone with a relatively **[choose option below]** unconscious bias compared to other people.

- lower (1)
  - on average (2)
  - higher (3)
  - Refuse to answer (4)
- 

Q2.6

I think my close colleagues have **[choose option below]** unconscious bias compared to me.

- a lower (1)
  - an equal (2)
  - a stronger (3)
  - Refuse to answer (4)
- 

Page Break

---

Q2.7 – Q2.9

*\*The Implicit Association Test (IAT) is a test developed by Harvard to measure a person's unconscious bias. A person can take this test for free online at [implicit.harvard.edu/implicit/takeatest.html](http://implicit.harvard.edu/implicit/takeatest.html). It takes up to 10 minutes.*

*\*\*ERG = Employee Resource Group. These groups have recently been created in Microsoft NL to raise awareness and tackle the negative stigma around D&I. Every employee of Microsoft NL can join these groups.*

To what extent do you agree with the following statements?

	Not at all (1)	Slightly (2)	Moderately (3)	Very (4)	Extremely (5)	Refuse to answer (6)
<p><b>1)</b> I will let others take the Implicit Association Test (IAT)* this week, because I find becoming aware of your unconscious bias in D&amp;I related topics important. (1)</p>	<input type="radio"/>					
<p><b>2)</b> I do not think it will help the business to make more people aware of their unconscious bias in D&amp;I related topics. (2)</p>	<input type="radio"/>					
<p><b>3a)</b> <i>If not in an ERG**:</i> I will immediately become a member of an ERG after this session. <b>3b)</b> <i>If being in an ERG**:</i> After this session, I will immediately share something interesting, fun, educational or motivation in the ERG Teams group of which I am part of. (3)</p>	<input type="radio"/>					

---

Page Break

End of Block: Questions

Start of Block: Controls after

Q3.1 What is your current gender identity?

- Male (1)
  - Female (2)
  - A different identity (3)
  - Refuse to answer (4)
- 

Q3.2 Are you Dutch?

- Yes (1)
  - No (2)
  - Refuse to answer (3)
- 

Q3.3 Age

- Under 18 (1)
  - 18 - 24 (2)
  - 25 - 34 (3)
  - 35 - 44 (4)
  - 45 - 54 (5)
  - 55 - 64 (6)
  - 65 or older (7)
  - Refuse to answer (8)
-

Q3.4 Do you have a mental or physical disability?

- Yes (1)
- No (2)
- Refuse to answer (3)

---

Page Break

Q3.5 Do you have a technical role at Microsoft?

- Yes (1)
- No (2)
- Refuse to answer (3)

Q3.6 Are you a manager?

- Yes (1)
- No (2)
- Refuse to answer (3)

Q3.7 What is your work duration at Microsoft NL?

0 6 12 18 24 30 36 42 48 54 60

---

Number or years (round up) ()



Q3.8 Are you part of an Employee Research Group (ERG)?

- Yes (1)
  - No (2)
  - Refuse to answer (3)
- 

Page Break

---

Q3.9 Have you taken an Implicit Association Test (IAT) before?

- Yes (1)
- No (2)
- Maybe / Do not remember (3)
- Refuse to answer (4)

End of Block: Controls after

---

Start of Block: Ending

Thank you so much for participating in this experiment! Please remain seated and silence until everyone is finished. This should not take too long.

If you are interested in my thesis, please do not hesitate to reach out to me (Isabelle) so I that I can update you.

**Do not forget to click one more time on the 'submit' button to submit your responses.**

End of Block: Ending

---