COMBATING DISCRIMINATION IN DATA ANALYTICS RECOMMENDATIONS

DEMOGRAPHIC TRANSPARENCY TO COMBAT
DISCRIMINATORY DATA ANALYTICS RECOMMENDATIONS

By SEPIDEH EBRAHIMI, M.Sc., B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial
Fulfilment of the Requirements for the Degree

Doctor of Philosophy in Business Administration

McMaster University DOCTOR OF PHILOSOPHY IN BUSINESS ADMINISTRATION (2018) Hamilton, Ontario, Canada

TITLE: Demographic Transparency to Combat Discriminatory Data Analytics Recommendations

AUTHOR: Sepideh Ebrahimi, M.Sc., B.Sc.

SUPERVISOR: Dr. Khaled Hassanein

NUMBER OF PAGES: x, 148

ABSTRACT

Data Analytics (DA) has been blamed for contributing to discriminatory managerial decisions in organizations. To date, most studies have focused on the technical antecedents of such discriminations. As a result, little is known about how to ameliorate the problem by focusing on the human aspects of decision making when using DA in organizational settings. This study represents an effort to address this gap. Drawing on the cognitive elaboration model of ethical decision-making, construal level theory, and the literature on moral intensity, this study investigates how the availability and the design of demographic transparency (a form of decisional guidance) can lower DA users' likelihood of agreement with discriminatory recommendations of DA tools. In addition, this study examines the role of user's mindfulness and organizational ethical culture on this process. In an experimental study users interact with a DA tool that provides them with a discriminatory recommendation. The results confirm that demographic transparency significantly impacts both recognition of the moral issue at hand and perceived proximity toward the subject of the decision, which in turn help decrease the likelihood of users' approval of the discriminatory recommendation. Moreover, the results suggest that user's mindfulness and organizational ethical culture enhance the positive impacts of demographic transparency.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor, Dr. Khaled Hassanein, who definitely contributed the most in my reaching this stage today and with whom I have learned what it really means to be a professor. Dr. Hassanein, I am convinced that you played a pivotal role in my finishing my Ph.D. During the past few years, not only did you steer me in the right direction with patience, dedication, and devotion but you also helped me feel calm and confident. Thank you for all your help and support and for being such a wonderful supervisor. I am looking forward to continuing our collaboration in the future.

I am also grateful to the other members of my supervisory committee, in particular Dr. Milena Head for her valuable suggestions, constructive feedback, and unprecedented optimistic attitude; and Dr. Yufei Yuan, for his challenging questions and critical insights that helped me improve my work.

I would also like to thank the staff at the DeGroote School of Business for their help that made my five-year journey easy and smooth. Deb Randall-Baldry and Kim Wilms, thank you for making the PhD office a supportive, safe, and fun place. Jeanette Hunter, I am grateful to you not only for your help throughout these years but also for the inspiring conversations.

Next, I would like to thank my colleagues and friends at DeGroote, who made this journey enjoyable and memorable. Special thanks go to Samira Farivar, who was my closest companion in this journey and who showed me that one's PhD cohort can be like a family, to Kamran Eshghi, whose brotherly friendship and support was so heartwarming, and to Maryam Ghasemaghaei with whom I shared many special moments in Hamilton, Dublin, Montreal, and Seoul.

A very big thanks goes to my family. Maryam, Davod, Sahar, and Ali, thank you for your loving support and understanding, without which none of this would have happened.

Last but certainly not least, I owe a huge thanks to my husband. Thank you Amirhossein for loving me and for showing so much understanding and support throughout this journey. Thank you for always being there for me especially in the final few months and whenever things seemed dark. Without you, I would not be writing this now.

Hamilton, ON, August 2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The last decade has witnessed a widespread adoption of computers, smartphones, and in general Internet-connected devices by organizations and consumers. This has enabled organizations to collect an ever-increasing amount of data, which they strive to analyze by employing data analytics tools in order to gain data-driven insights and make data-driven decisions (Ghasemaghaei et al. 2018). Data Analytics (DA) is the process of analyzing large amounts of data using computer systems to discover patterns in support of decision making (Shang et al. 2013). Data analytics is often a combination of a number of processes and tools, including SQL queries, statistical analysis, data mining, fact clustering, and data visualization and is a way to discover customer segments, associate similar and related products, etc. (Russom 2011). Such an approach to decision making has been suggested to be superior to 'HIPPO' (highest-paid person's opinion) style (McAfee and Brynjolfsson 2012). As a result, data-driven decision making not only contributes to higher financial and strategic performance in organizations (Côrte-Real et al. 2017) but also can lead to governments serving their citizens better, hospitals being safer, law enforcers catching more criminals, etc. (Martin 2015). Nonetheless, data analytics has been accused of contributing to privacy breaches as well as discrimination in societies (Newell and Marabelli 2014). This thesis focuses on the latter issue (i.e., potential discrimination as a result of data analytics use) and strives to propose a method to help reduce the incidence of such discrimination.

Discrimination occurs when member(s) of a socially defined group are treated differently (especially unfairly) because of their membership of that group (Krieger 1999). The notion of discrimination as studied by social psychologists is often accompanied by concepts such as prejudice and stereotyping (e.g., Fiske 2000). Whereas it is suggested that technology has been employed in order to mitigate the problem of discriminatory decisions made based on personal prejudice (Gates et al. 2002; Tene and Polonetsky 2013), the issue has not been resolved by using data analytics tools to support decision making in organizations. On the contrary, it has been argued that using data analytics tools and techniques can contribute to discrimination in societies (Danna and Gandy 2002; Johnson 2014; Lyon 2003; Newell and Marabelli 2015). Even fair and well-intentioned decision makers can make discriminatory decisions using data analytics tools due to biased or non-representative data as well as inadvertent modeling procedures in the DA tools they use to support their decision making (Žliobaitė and Custers 2016). There are three main reasons as to why unbiased algorithms might generate potentially discriminatory outcomes: (i) relations between non-sensitive and sensitive attributes in data; (ii) data labeling; and (iii) data collection (Calders and Žliobaitė 2013)[1]. O'Neil (2016) in her book on "Weapons of Math Destruction" states this fact in a nutshell: "The computer learned from the humans how to discriminate, and it carried out this work with breathtaking efficiency" (p. 116).

[1] These reasons will be discussed in more details in Chapter 2.

1.1 Research Motivation

As discussed above, the problem of unethical discriminatory recommendations by data analytics systems, often arises due to reasons other than basing the recommendation on sensitive variables (e.g., sex, age, and race). However, although the system and the analyst might not include those sensitive variables as inputs into the analysis, the abundance of data on individuals being analyzed, biased or non-representative data, and inadvertent modeling procedures can lead to generation of discriminatory outcomes (i.e., different outcomes based on individuals' protected class membership) (Žliobaitė and Custers 2016). Such discriminatory outcomes have been called disparate impact, adverse effect, and indirect discrimination[2]. Making decisions that includes adverse effect is not only discriminatory and thus unethical but is also illegal in many parts of the world (For a review of the related laws in different countries and International laws, see Hunter and Shoben 1998).

While, to the best of my knowledge, no rigid mathematical formula exists to frame disparate impact, there have been some suggestions as to what constitutes an instance of disparate impact. A brief review of these measures will be provided in Chapter 2. For instance, the US Equal Employment Opportunity Commission, the civil Service Commission, the Department of Labor, and the Department of Justice have jointly adopted a rule, known as "the four-fifths rule" to identify such instances. The four-fifth

[2] It should be noted that disparate impact (aka adverse effect and indirect discrimination) is different from disparate treatment (aka direct discrimination), which refers to discriminatory behavior toward someone(s) due to their membership in a protected class.

rule states that a selection rate for any race, sex, or ethnic group, which is less than four-fifths (i.e., 80%) of the selection rate for the group with the highest selection rate will be regarded as evidence of disparate impact (Feldman et al. 2015). However, such a rule is neither practiced in the US nor globally in a way that a violation of it is considered as illegal. As a result, this study does not focus on forcing any rule to be practiced; instead, it only draws on the notion of disparate impact and the various selection rates in different demographic groups.

While some technical methods have already been suggested for discovering and removing discrimination in data mining procedures (e.g., Dwork et al. 2012; Kamiran et al. 2010; Pedreshi et al. 2008), developing computational means to prevent such discrimination is an ongoing endeavor (Žliobaitė and Custers 2016). The insufficiency of the existing methods to eliminate the aforementioned discrimination is evident in recent scholarly and practitioners' (Crawford 2013; McDonald 2016; Newell and Marabelli 2015; Schrage 2014) and even governments' (Federal Trade Commission 2016; Podesta et al. 2014) publications raising concerns and awareness about the potential of making discriminatory decisions using data analytics tools.

In organizations, it is ultimately the managers and decision makers' responsibility to make sure that their data-driven decisions are free of discrimination. However, algorithms (e.g., Google search algorithm) are often multi-component systems built by teams and therefore, include some level of opacity that even the programmers who are insiders to the algorithms must deal with (Sandvig et al. 2014). In the case of machine learning algorithms, the opacity is even higher since the internal decision logic of the algorithm is altered as it learns on training data (Burrell 2016). As a result, in organizations, few individuals actually understand the algorithms included in data analytics tools (Newell and Marabelli 2015). Therefore, often finding a strong predictive association by an algorithm is seen as sufficient and finding out the reasons for those associations in the data from different sources are neglected (Newell and Marabelli 2015). The fact that barely anyone in an organization knows why some decisions are made can cause issues including the unethical issue of making discriminatory decisions.

It is important to note that fairness issues can arise after the model has been built and while it is be deployed. Indeed, the extent of differential impacts of a model on different groups mainly becomes evident once it is used in a decision making system; for example, the impact of the setting of thresholds for positive and negative outcomes could have significant consequences for different groups that cannot be evident by studying the model itself (Veale and Binns 2017).

As a result, discriminatory decisions are best identified through investigating their outcomes. In other words, while it might be difficult to investigate the process through which a discriminatory decision has been made, it is relatively easier to assess the outcome of an analysis/decision to identify traces of discrimination. The complexity of investigating the process of a discriminatory recommendation being put forth by data analytics is challenging and sometimes even impossible as few individuals actually understand what is included in the algorithms and why (Newell and Marabelli 2015). As such, to help decision makers using data analytics tools to recognize traces of discrimination (if any) in the system's recommendation, this study suggests providing them with some individual and aggregated demographic information as a form of decisional guidance.

More specifically, this study suggests providing DA users with demographic information that shows the proportion of members of each demographic class in both the original pool and the recommended sample by the data analytics tool to enable the users to compare and investigate these proportions in regard to any possible discrimination in the recommendation. This suggestion is in line with the common practices in identifying the discriminatory decisions in the business context such as the four-fifth-rule as well as other measures used in the literature, that will be discussed in more details in Chapter 2. It is important to note that as per the statement on "Algorithmic Transparency and Accountability" issued by the Association for Computing Machinery, organizations "should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results" (Dopplick 2017). As such, this study suggests that providing DA users with the above-described demographic information equips them with an aid and guidance to identify recommendations generated by DA tools that might be discriminatory. It is noteworthy that there is no guarantee that DA users will act on the demographic information provided to them. Nonetheless, the hope is that the provision of such information will reduce the incidences of their readily accepting potentially discriminatory recommendations generated by DA tools.

1.2 Research Objectives

The objectives of this study are two-fold. As the first objective, the impact of providing Data Analytics users with aggregated and individual demographic information about the subjects of the analysis and decision on their acceptance of a potentially discriminatory recommendation generated by the DA tool will be investigated. In other words, as it is difficult to prevent data analytics from generating discriminatory recommendations as well as it is difficult if not impossible to understand the process through which the recommendations have been put forth, this study focuses on providing data analytics users with decisional guidance to help them scrutinize the recommendations put forth by a DA tool. More specifically, this study examines the impact of providing the decision maker with individual and aggregated data regarding the demography of human subjects of the data analytics recommendations based on the original data set and the recommended sample and examines whether such data can help decrease the likelihood of approving a potentially discriminatory recommendation provided by a data analytics tool.

Providing DA users with demographic information is in line with recent conceptual and technical studies that argue collecting sensitive personal data and using them in the analyses can have significant impact on reducing the likelihood of making unethical decisions using DA tools (Williams et al. 2018; Žliobaitė and Custers 2016). Žliobaitė and Custers (2016) using standard regression models show that including sensitive information (e.g., race) in the model building process decreases the likelihood of building a discriminatory decision model. Williams et al. (2018) discuss how using data analytics can amplify and reinforce the societal biases. They demonstrate how a lack of social categories can exacerbate existing biases by making them harder to detect and addresses. Therefore, they argue that organizations should collect and carefully use social category data in their data analysis using DA tools.

In line of the above discussion, the first research objective of this study is:

- *To employ a theoretical model to investigate how and to what extent does providing aggregated and individual demographic information regarding the human subjects of DA recommendations would reduce the incidence of users' acceptance of potentially discriminatory recommendations of DA systems.*

Ethical/unethical actions take place in social contexts and by individuals with various characteristics. Therefore, while studying ethical/unethical behaviors, it is important to take into account individual as well as situational variables. Linda Trevino in her seminal paper, suggests a person-situation interactionist model of ethical decision-making and contends "ethical/unethical behavior in practical situations […] results from an interaction between the individual and situation" (Trevino 1986, p. 610). Trevino further breaks down the situational variables to the ones arising from the immediate job context and the broader organizational culture.

Likewise, individual variables and organizational culture are expected to affect the impact of providing DA users with aggregated and individual demographic information on their acceptance of a potentially discriminatory recommendation put forth by the DA tool. The specific individual characteristic of interest in this study is mindfulness, which has received attention in both the IS (e.g., Butler and Gray 2006) and ethics (e.g., Ruedy and Schweitzer 2010) literatures. Mindfulness defined by Brown and Ryan (2003, p. 822) as “a state of being attentive to and aware of what is taking place in the present” can potentially impact the success of provision of aggregated and individual demographic information in reducing the likelihood of accepting potentially discriminatory recommendations of data analytics tools.

As for the organizational characteristics and culture, the pivotal role of organizational ethical culture has been underlined in many studies (e.g., Rottig et al. 2011; Sweeney et al. 2010). Indeed culture is capable of influencing individuals’ beliefs and behavior (Ferrell and Gresham 1985; Hunt and Vitell 1986). In a review of business ethics literature, Loe et al. (2000, p. 194) emphasize the important role of an ethical culture in organizations and state that “Perhaps our greatest opportunities for research relate to evaluating the effectiveness of codes, structuring of codes, and their communication and integration with other aspects of the organization’s culture”. Therefore, the second research question of this study is:

- *To investigate how and to what extent do user’s mindfulness and organizational ethical culture impact the relationship between providing aggregated and individual demographic information regarding the human subjects of DA recommendation and DA user’s acceptance of potentially discriminatory recommendations of those systems.*

1.3 Dissertation Contributions and Structure

This study makes three major contributions to the literature. Firstly, it illuminates the effect of providing aggregated and individual demographic information on data analytics users' assessment/acceptance toward DA recommendations that could be discriminatory. Secondly, it proposes and operationalizes a method to provide data analytics users with aggregated and demographic information about the subjects of their analyses that could help them identify potentially discriminatory DA tools' recommendations. Thirdly, it assesses the effect of users' mindfulness and organizational ethical culture on the effectiveness of providing such aggregated and individual demographic information in regard to reducing the likelihood of users' accepting potentially discriminatory recommendations. In addition, this study could have strong implications to the practice of using data analytics tools in organization and how to best support DA users.

The rest of this dissertation is structured as follows. Chapter 2 provides a contextual background about the discriminatory recommendations of data analytics tools and the measures used in the literature to measure discrimination. Subsequently, Chapter 3 presents the theoretical background for this research and details the proposed theoretical model and associated hypotheses. Chapter 4 describes the experimental methodology for collecting data to empirically validate the proposed model. Chapter 5 delineates the preliminary data analyses and the results of statistical tests of our hypotheses. Finally, Chapter 6 outlines the study's contributions to theory and its implications for practice as well as its limitations and avenues for future research.

2. CONTEXTUAL BACKGROUND

Discriminatory decision-making and behavior is certainly not new and has been observed many years before deployment of data mining and data analytics for decision-making. The concept of discrimination has been given several meanings and definitions. Nonetheless, for the purpose of this study the following definition by Andrea Romei and his colleagues is adopted as similar definitions have been used in studying discrimination in the context of data mining and data analytics. As such, "Discrimination refers to an unjustified distinction of individuals based on their membership, or perceived membership, in a certain group or category disregarding individual merits" (Romei et al. 2013, p. 6064). Discrimination is mainly categorized into direct discrimination and indirect discrimination. Direct discrimination refers to a situation where an individual is treated adversely directly on the basis of prohibited grounds such as when an employer categorically refuses to hire immigrants (Makkonen 2002). Direct discrimination has been also called disparate treatment (Gardner 1989). Indirect discrimination, on the other hand, does not discriminate against individuals directly based on their membership in protected groups. However, it consists of rules, procedures or requirements that without explicitly mentioning discriminatory attributes, impose disproportionate burdens on individuals from protected groups (Pedreshi et al. 2008). Indirect discrimination is also called disparate impact (Collins 2003).

In order to fulfill the main objective of this study, which is investigating whether providing aggregated and individual demographic information regarding the human subjects of DA recommendations would reduce the incidence of users' acceptance of potentially discriminatory recommendations of such tools, it is required to first have an understanding of why such recommendations can be generated by DA tools. In addition, it is critical to learn what measures are most used to quantify and measure discrimination in the literature and by legislation. The following two sections discuss these two topics.

2.1 Reasons behind Generation of Potential Discriminatory Recommendations by Data Analytics Tools

One main method used extensively by data analytics tools is profiling, which can be performed through several different processes such as classification and clustering[3]. Profiling has been associated with generating discriminatory recommendations by data analytics tools (Danna and Gandy 2002; Newell and Marabelli 2014). Custers (2004) in his book on ethical, legal, and technological aspects of data mining and group profiling in epidemiology describes the reasons for problems arising from group profiling using technology. He discusses the problem of masking, which occurs when trivial information are correlated with sensitive information (e.g., when people living in a certain zip code have a high health risk and insurance companies use zip code as a selection criterion). In addition, Custers asserts that data mining approaches investigate every possible relation

[3] It is noteworthy that profiles can be divided into two main categories. Whereas, an individual profile is a collection of properties of a particular individual, a group profile is a collection of properties of a particular group of people. The category of interest in this study is group profiling as the problems that arise from individual profiling is mainly related to the heading of 'privacy'.

and, therefore, are different from the causal relationships considered in empirical statistical research (Custers 2004). This can lead to the problem of illusory correlations (i.e., when statistical relations are interpreted as causal relations). Furthermore, data mining approaches may lack transparency and, therefore, it is harder for individuals to protect themselves against group profiling. For example, one might be reluctant to share their sensitive information with an organization; however, they might be unaware of the fact that by providing their trivial information to the same company, they can be categorized into a group profile of people about whom sensitive information is known (Custers 2004).

It should be noted, however, that not all unethical discriminatory decisions are made by individuals with vicious intentions. Calders and Žliobaitė (2013) discuss three main reasons for unbiased data mining algorithms generating discriminatory outcomes: (i) relations between non-sensitive and sensitive attributes in data; (ii) data labeling; and (iii) data collection. These are discussed below:

(i) The attributes that characterize the subject might not be independent from each other and as a result, (neutral) factors relevant to a rational decision making process (e.g., zip codes) might act as “proxies” for sensitive variables (e.g., race). Therefore, simply dropping the sensitive variables from the analysis does not guarantee nondiscriminatory results. One famous example of using a neutral variable resulting in discriminatory decisions is the redlining practice (See Massey and Denton 1993). Redlining is the practice of arbitrarily denying products and services in specific neighborhoods, marked with a red line on a map to show where not to provide such services. This practice often resulted in discrimination against people of color (Hunt 2005).

As another example, consider organizational contexts, where certain criteria are used to make employment and promotion-related decisions. However, there can be strong relationships between those criteria and membership in protected classes. Therefore, employers, while providing greater opportunities to employees that they predict will outperform others at some task, might find that they treat members of some protected groups disadvantageously because the criteria used to determine the attractiveness of employees happen to be systematically held at lower rates by members of these groups (Barocas and Selbst 2016; Kamiran et al. 2013). In such a case, a decision maker does not maliciously discriminate due to holding prejudicial beliefs, but he/she unintentionally makes a discriminatory decision that repeats the inequality that exists in the society.

(ii) The historical data in organizations contain labels based on which data mining models are built and trained. As such, discriminatory data can lead to discriminatory models (Custers 2013) because if cases in which prejudice has played a role are treated as valid examples from which a decision-making rule is learned, chances are high that the rule reproduces the prejudice involved in the earlier cases. Example of subjective labels that might include traces of discrimination are the decision to whether or not to detain a suspect, assessment of a human resource manager of whether a job candidate is suitable for a specific job, etc. (Calders and Žliobaitė 2013). As another example consider websites such as LinkedIn Talent Match system that provide employers with recommendations about potential employees (Woods 2011). If LinkedIn determines its recommendations based on the demonstrated interest of the previous/current decision makers in organizations, Talent Match will offer recommendations that recapitulate the biases (if any) that employers have shown. The same is true when in defining a *good employee*, an employer takes into account parameters annual review grades, which might bring in the personal prejudice of previous managers into the analysis process (Stauffer and Buckley 2005).

(iii) The data collection process may be intentionally or unintentionally biased. Problems due to data collection occurs when some groups of individuals are over- or under-represented in the training data set[4] of data analytics algorithms. For example, when police officers, due to their personal prejudice, have targeted an ethnic minority disproportionately in the past, chances are high that the same ethnic minority would appear more prominently in crime statistics. Consequently, if such data is used to train a classifier, it is likely that the classifier will learn that a strong correlation exists between ethnicity and crime (O'Neil 2016; Schermer 2011). The same concept is also applicable to the context of managers who devote disproportionate scrutiny to workers of certain

[4] Training data is the data set that is used to build the model.

> protected classes (e.g., race, sex) and subsequently find and document more instances of mistakes in their work. Such a data set, if used as training data to build analytical models, might lead to generating discriminatory outcomes for members of those protected classes (Barocas and Selbst 2016).

Another reason for discrimination discussed by some scholars is when the full data set[5] being analyzed is biased (unbalanced) itself and does not proportionally represent the entire population. This might happen due to reasons such as some protected classes' having less access to the Internet and other high tech technologies, for example (Barocas and Selbst 2016). In addition, such a problem can arise due to members of a protected class being treated unfairly by some algorithms. For instance, recent research has demonstrated that setting the gender to female results in receiving fewer instances of an ad from Google relating to high paying jobs than setting it to male (Datta et al. 2015). Therefore, it is likely that male applicants outnumber female applicants in the pool of applications for a high paying job.

In light of the above discussion, the discrepancies in the data analysis can be grouped into four categories (see Figure 2.1). Category 1 includes instances where the full data set includes a disproportionate representation of a particular class (e.g., male applicants for a job, black criminals). In such a situation chances are high that the recommended sample (to be invited for an interview or to be scrutinized for crimes, for example) will too be biased in favor of/against the demographic class that is dominant in the full data set.

[5] It is notable that this data set is different from the training data set. In this stage, the model is already built and is being applied on a new data set.

Category 2 represents instances where the full data set is balanced; yet, the recommended sample is unbalanced as it includes disproportionate increases in the share of a demographic class compared to the original full data set. Such an instance can happen due to reasons discussed earlier (e.g., biased training data and neutral variables acting as proxies for sensitive variables). Category 3 represents instances where the full data set is unbalanced; however, the recommended sample is not. Such a situation might occur if a DA tool includes some features (e.g., amended weights of factors, for example) to prevent discrimination. Finally, category 4 includes instances that both the full data set and the recommended sample are balanced.

As is evident from the above discussion, while categories 1 and 2 include potential discrimination against one or more demographic group/s, categories 3 and 4 are most probably free of discrimination. In addition, whereas category 4 can be considered as an ideal situation, category 3 can be potentially troublesome because if the full data set remains unbalanced and biased, it might bring about discrimination-related issues in the future involving other analyses with this data set.

Category 1 Full data set: Unbalanced Recommended sample: Unbalanced	**Category 2** Full data set: Balanced Recommended sample: Unbalanced
Category 3 Full data set: Unbalanced Recommended sample: Balanced	**Category 4** Full data set: Balanced Recommended sample: Balanced

Figure 2.1. Categories of Possible Discrepancies in the Data

2.2 Discrimination Measures

Measuring is pivotal in discovering and preventing discrimination in algorithmic decision-making (Pedreschi et al. 2009). Thus, a brief overview of the main used discrimination measures is provided in this section. There are also a few related reviews in the literature that avid readers can refer to (e.g. Romei and Ruggieri 2014; Zliobaite 2015). Let us first review the notation that will be used in this section: v denotes the protected variable (e.g., sex). y represents the target variable, which shows the decision (class) item. As in investigating possible discrimination, one of the main goals is to analyze the cases which belong to the protected group yet receive the desired decision, here v=1 refers to the protected community and y=1 refers to the desired decision (e.g., positive decision to grant a loan). Table 2.1 depicts more details about the notation employed in this section.

One discrimination measure used in several studies is called extended lift. Extended lift (elift) which measures how the probability of granting a benefit to an individual changes as a result of their belonging to a protected group (Pedreshi et al. 2008). The Extended lift measure is calculated as the (Zliobaite 2015):

$$elift = \frac{p(y = 1|v = 1)}{p(y = 1)}$$

Table 2.1. Summary of Symbols Used	
Symbol	**Explanation**
y	Target variable (typically a binary variable, where y=1 denotes positive desirable outcome and y=0 denotes negative outcome)
v	Protected variable, $v \in \{1,2,\ldots,m\}$ v=1 denotes a protected group
x	A set of input variables (predictors)
n_i	Number of individuals in group i

For instance, a rule SEX=FEMALE, CITY=NYC → HIRE=NO with an extended lift of $\frac{1}{3}$ means that compared to average people living in New York City, being a female increases 3 times the probability of not being hired. In other words, the above rule is compared to CITY=NYC → HIRE=NO and the results show that the probability of the former is three times less than the probability of the latter. In such a scenario women are potentially discriminated against[6]. It is noteworthy that below is the alternative yet equivalent way to defining the extended lift:

[6] It is noteworthy that as mentioned previously, this is a case of potential discrimination as one might argue that there are legitimate business necessities behind such a situation (e.g., when driving a truck is a requirement for the job and fewer females have a license to drive a truck).

$$elift = \frac{p(v = 1|y = 1)}{p(v = 1)}$$

The above formula makes it clear that extended lift is related to the principle of group over-representation in denying a benefit (or equivalently group under-representation in granting a benefit) (Pedreschi et al. 2013) as discussed in the previous section. The notion of extended lift is also in line with non-discrimination law in several countries that state that individuals should not be "differentiated adversely […] on a prohibited ground of discrimination" (Canadian Department of Justice 2017) or be treated "less favourably" than others because of possessing a protected characteristic (European Union Legislation 2011; U.K. Legislation 2011) and that a qualifying criterion should not be defined in a way that "a higher proportion of people without the [protected] attribute comply or are able to comply" (Australian Legislation 2015).

There are also other measures to identify traces of potential discrimination in a situation. For instance, selection lift (slift) or impact ratio compares the outcome (i.e., target variable) for individuals with a protected attribute (e.g., black people) with the outcome for individuals without the protected attribute as the following (Zliobaite 2015):

$$slift = \frac{p(y = 1|v = 1)}{p(y = 1|v \neq 1)}$$

Similarly, contrasted lift (clift) is defined by the ratio that compares the outcome for individuals with a protected attribute (e.g., black people) with the most advantaged group (e.g., white people) as:

$$clift = \frac{p(y = 1|v = 1)}{p(y = 1|v = 2)}$$

Where v=1 refers to black individuals and v=2 refers to white individuals in the dataset. The $clift$ measure is in line with what is known as the four-fifth rule that is being used by some organizations in the US (e.g. US Equal Employment Opportunity Commission) for quantifying discrimination. The four-fifth rule states that a selection rate for any race, sex, or ethnic group which is less than four-fifth (or eighty percent) of the rate for the group with the highest selection rate among all groups will generally be regarded as evidence of adverse impact (Equal Employment Opportunity Coordinating Council 1976). Therefore, the four-fifth rule has determined a maximum threshold in the US for contexts such as employment decisions.

Drawing on the natural and biomedical sciences, Odds ratio ($olift$) as another ratio measure for quantifying discrimination has been suggested. Odds ratio instead of the probability of granting a benefit (i.e., p), considers the odds of such an event (i.e., p/(1-p)). For instance in a hiring situation, the odds ratio is the ratio between the odds of hiring a person belonging to a protected group over the odds of hiring a person not belonging to that group (Pedreschi et al. 2009).

$$olift = \frac{\frac{p(y = 1|v = 1)}{p(y = 0|v = 1)}}{\frac{p(y = 1|v \neq 1)}{p(y = 0|v \neq 1)}} = \frac{p(y = 1|v = 1)p(y = 0|v \neq 1)}{p(y = 1|v \neq 1)p(y = 0|v = 1)}$$

Zliobaite (2015) compares performances of extended lift, selection lift, and odds ratio on synthetically generated data that were changed to represent different task settings and different levels of underlying discrimination. She concludes that although not perfect, the extended lift measure outperforms the selection lift and odds ratio measure as it is less sensitive to and more stable in cases of imbalances in the original data set (e.g., more males among the applicants for a job than females).

The measures introduced thus far are defined in terms of ratios; however, some measures based on the difference of selection probabilities have also been introduced and considered. More specifically, difference measures mainly rely on how the proportions of benefits granted for the protected group (e.g., females) is different from the one for the unprotected group (e.g., males). For instance, consider the following formula called discrimination score (Calders and Verwer 2010) as well as risk difference (RD) and absolute risk (Romei and Ruggieri 2014).

$$RD = p(y = 1|v = 1) - p(y = 1|v \neq 1)$$

Similar to the notion of extended lift introduced at the beginning of this section, extended difference (*eRD*) is defined as (Pedreschi et al. 2009; Pedreschi et al. 2013):

$$eRD = p(y = 1|v = 1) - p(y = 1)$$

As the above discussion shows, there is no general agreement on a standard measure for discrimination neither in the literature nor by legislations. Nonetheless, a general principle is to consider group under-representation as a quantitative measure that people in a group have been treated less favourably (Pedreschi et al. 2009). Drawing upon this understanding of quantitative measures of discrimination, the next chapter will propose and delineate the notion of demographic transparency that aims at providing data analytics users with aggregated and individual demographic information in order to enable them to scrutinize the potential discriminatory recommendation of these tools.

3. THEORY DEVELOPMENT

As discussed in Chapter 1, the main objective of this study is to assess the effectiveness of the provision of aggregated and individual demographic information about the human subjects of data analytics recommendations on reducing users' acceptance of potentially discriminatory recommendations put forth by those systems. To that end, this study draws upon the cognitive elaboration model of ethical decision-making and the literature on moral intensity, which will be discussed in the following sections.

3.1. Cognitive elaboration model of ethical decision-making

The cognitive elaboration model of ethical decision-making enables us to gain a better understanding of how providing extra information in the forms of aggregated and individual demographic information about the subjects of the DA analyses can decrease the likelihood of the users' acceptance of potentially discriminatory DA recommendations. The cognitive elaboration model of ethical decision-making draws upon two major theories in the psychology literature: Rest's (1986) four-component model for individual ethical decision-making and behavior as well as the Elaboration likelihood model (ELM) (Petty and Cacioppo 1986b) to explain the impact of cognitive expenditure on the ethical decision-making process.

Rest's (1986) four-component model of ethical decision making is undoubtedly one of the most prevalent models in the ethics and business ethics literatures. Rest argues that, during the course of making a decision involving an ethical dimension, individuals move through a series of four sequentially ordered steps, namely, recognition of the moral issue, making a moral judgment, establishing the intent to act morally, and engaging in a moral behavior. The first step, recognition of the moral issue, also known as moral awareness, is an interpretive process in which the individual recognizes that a moral problem exists in a situation or that a moral principle is relevant to the existing set of circumstances (Reynolds 2006). Recognition of the moral issue, then, prompts the decision maker to make a judgment of what potential action is most moral. Moral intent is prioritizing moral values over other values and finally moral behavior is the application of the moral intent to the situation (Craft 2013). It is important to note that recognition of a moral issue plays a pivotal role in the process of making an ethical decision as without recognizing the moral issue the process might not be triggered at all as a person who fails to recognize the moral aspect of an issue will fail to employ ethical decision-making schemata and will make the decision based on other schemata such as economic factors, etc. (Jones 1991).

Recognizing the moral aspect of an ethically laden issue (i.e., moral awareness) is a form of attitude change toward an object/event (Street et al. 2001). The ELM (Petty and Wegener 1998; Petty and Cacioppo 1986a) provides a theoretical perspective on how individuals' attitudes toward objects/events are formed in different circumstances. The ELM argues that contingent on the extent of content evaluation and issue-relevant reflection (i.e., elaboration), attitude change can occur through two main routes: central route and peripheral route. The central route is invoked when an individual engages in diligent and effortful information processing regarding a decision or event (Petty et al. 1995). However, since scrutinizing information consumes cognitive resources, the central route is not activated in all circumstances. When the individual is not motivated and/or does not have enough ability to engage in evaluating a decision, object, or event, he/she exerts minimal cognitive effort, follows the peripheral route, and evaluates the information based on some peripheral cues associated with the information and not the information itself (Krosnick and Petty 1995).

Attitudes can change "from no attitude to some attitude […], or from one attitudinal position to another" (Petty et al. 1995, p. 94). In the context of an ethical decision, when an individual faces an ethically charged issue, they can hold one of three attitudes regarding the issue: they can have no opinion at all about the issue; or they can incorrectly believe that there is no moral aspect to the situation; or they can hold the correct view that the issue does have moral aspects. The first two attitudes result in the individual failing to recognize the ethical ramifications of the issue, which leads the decision to be made based on non-moral considerations (Street et al. 2001). Holding such attitudes can be the case in many situations because the ethical ramifications of the issue is not always blatantly obvious and not all moral issues come with a waving red flag that says "Hey, I'm an ethical issue. Think about me in moral terms!" (Trevino and Brown 2004, p. 70). As a result, as suggested by cognitive elaboration model of ethical decision-making, when an individual faces an issue with ethical aspects, if the overall influence of the available ability (e.g., relevant knowledge) and motivation (e.g., personal relevance) factors results in a high level of elaboration, the individual is more likely to use their central information processing route and, therefore, they are more likely to recognize the moral aspects of the decision at hand (Street et al. 2001). In other words, the decision maker should have the required ability and motivation to scrutinize the information and the context regarding the decision in order to recognize the ethical aspects of the moral issue at hand.

3.2. Moral Intensity

One important motivational variable discussed in the cognitive elaboration model of ethical decision-making is the notion of moral intensity introduced by Jones (1991). Using the cognitive psychology literature, Jones suggests that recognition of moral issues is due to their salience and vividness (May and Pauli 2002). The salience of a stimulus is the extent to which it stands out from its background. The vividness of a stimulus is the extent that it is "(a) emotionally interesting, (b) concrete and imagery provoking, and (c) proximate in a sensory, temporal, or spatial way" (Nisbett and Ross 1980, p. 45). Issues with higher levels of salience and vividness are more likely to be recognized as moral issues since due to their nature tend to dominate people's attention.

Jones (1991) attributes such salience and vividness to the *moral intensity* of the issue, which is comprised of six issue-contingent factors. First, *magnitude of consequences* suggests that the issue will be more serious if its sum of harms (or benefits) done to the victims (beneficiaries) is higher. For instance, an action that leads to death of one person is of higher magnitude of consequences compared to an action that causes a minor injury to one individual. Second, *social consensus* states that a higher level of social agreement that a proposed behavior is evil (or good) makes the issue more intense. Jones suggests that including social consensus as a dimension of moral intensity has logical reasons as it is difficult to act ethically without knowing what good ethics prescribes in a specific situation. Third, *probability of effect* suggests that an issue will be more intense if it has higher likelihoods to occur and to cause the anticipated harm. Similar to the context of a financial gain, where expected value is defined as the product of the magnitude of the sum of the gain and its probability of occurrence, in the context of a harmful decision, the consequence of the decision would be the product of the magnitude of consequences, the probability that the harmful act will take place, and the probability that the act will cause the harm (Jones 1991). Fourth, *temporal immediacy* suggests that an issue with a shorter interval between when the decision is made and when the consequences occur is perceived as being more intense. Temporal immediacy is a dimension of the moral intensity of an issue due to the fact that people tend to discount the probability and the impact of events that happen in the future. Fifth, *proximity* states that the feeling of closeness that the decision-maker has for victims (beneficiaries) makes the issue more intense. For instance, layoffs in one's work unit has a higher moral proximity compared to layoffs in a remote plant. Finally, *concentration of effect* suggests that an issue is perceived as being more intense if the consequences affects fewer individuals as opposed to the same consequences being more broadly distributed (Jones 1991). The concept of concentration of effect is in line with the philosophy of ethical utilitarianism, which holds that "an act is right only if it produces for all people a greater balance of good consequences over bad consequences than other available alternatives (i.e., 'the greatest good for the greatest number')" (Hunt and Vitell 1986, p. 7). Therefore, an act that is "extremely bad" for a few people has a higher concentration of negative effect and consequently has a higher moral intensity than another act that is "bad" for a large number of people (Singhapakdi et al. 1996b). These dimensions of moral intensity, as predicted by Jones, have been widely shown to have a high impact on the ethical decision making process (e.g., Butterfield et al. 2000; Leitsch 2004; Singhapakdi et al. 1999; Singhapakdi et al. 1996b; Valentine and Bateman 2011).

3.3. Research Model and Hypotheses Development

To fulfill the research objectives outlined earlier, this study draws upon the above theoretical foundations, to propose the research model (depicted in Figure 3.1) and seven associated hypotheses, detailed below.

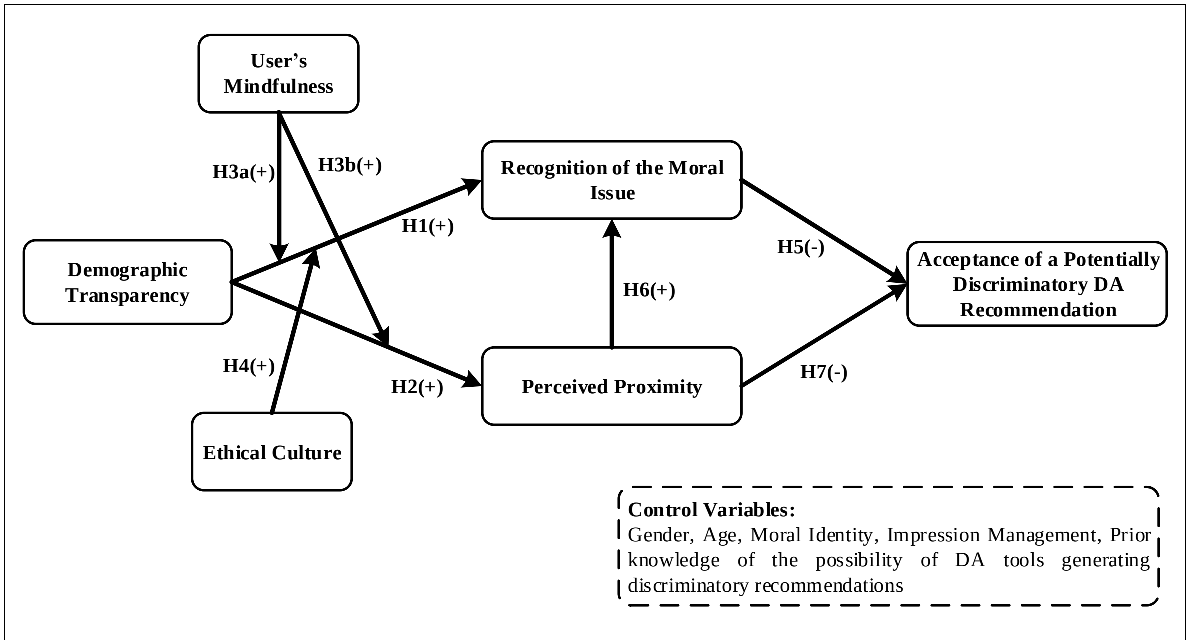


Figure 3.1. Research Model

3.3.1. Demographic Transparency

As discussed in the first two chapters, discriminatory recommendations of data analytics systems, are often generated due to reasons beyond using sensitive variables like race as inputs in the analysis. Nonetheless, due to the abundance of data on individuals being analyzed, biased or non-representative data, and inadvertent modeling procedures, discriminatory recommendation can be generated that are in favor/against a specific demographic group (i.e., different outcomes based on individuals' protected class membership) (Žliobaitė and Custers 2016). Such discriminatory outcomes are called indirect discrimination or disparate impact.

To establish presence of potential discrimination in the context of a selection system, substantial statistical disparity must exist between those from a protected class (e.g., women) and those from the most well-represented group (e.g., men) (Arakawa and Sonen 2009). As discussed in chapter 2, there have been a few metrics developed to measure whether a decision includes potential discrimination. One such statistic is extended lift that discusses how the probability of granting a benefit to an individual changes as a result of their belonging to a protected group (Pedreshi et al. 2008). More specifically, extended lift is defined as the ratio between the proportion of individuals from a protected class (e.g., females) obtaining a benefit over the overall proportion of them in the data set. Therefore, extended lift identifies whether a protected group is under-represented in a group receiving a benefit (or over-represented in a group denied a benefit).

Drawing on the notions of indirect discrimination and extended lift, and due to the fact that it is difficult, if not impossible, to investigate the process through which recommendations are generated by data analytics tools, this study recommends and investigates the effects of providing decision makers with demographic information about the human subjects of their decision. Such statistics aim at providing DA users with the proportion of members of each demographic class in both the original pool of subjects and the recommended sample by the data analytics tool.

Such demographic information is a form of decisional guidance defined by Silver (1991, p. 107) as “how a decision support system enlightens or sways its users as they structure and execute their decision-making processes”. Decisional guidance do not necessarily aim at steering decision makers in a given direction (Wang and Benbasat 2007) but might intend to provide 'meta-support' for the decision maker's judgmental activity (Silver 1991). Decisional guidance can take one of the two major types: (1) suggestive: making judgmental recommendations (e.g., what to do) to the decision-maker, and (2) informative: providing the user with unbiased and pertinent information without suggesting how to act (Silver 1991). In order to deal with the aforementioned disparate impact (i.e., indirect discrimination), the particular form of decisional guidance to be examined in this study is *demographic transparency*. It is defined as informative guidance aimed at helping users compare the proportion of each demographic class (e.g., female and male) in its pertinent demographic category (e.g., sex) in the original dataset and the DA recommended sample. For instance, when analyzing data about internal applicants for a position within an organization, demographic transparency will provide the user with the proportion of female and male applicants in both the original pool of all applicants and the recommended sample of applicants to be further interviewed. For instance, imagine a situation, where the original pool includes 100 applicants from which 47 (47%) are female and 53 (53%) are male. The system's recommendation will be judged to be potentially discriminatory if its recommended sample of 25 applicants to be invited for an interview includes 4 (16%) and 21 (84%) females and males respectively. This is because the proportion of male (female) applicants who were included in the recommended sample is considerably more (less) than the proportion of male (female) applicants in the original pool. This study suggests that without providing a user with such aggregated demographic information about the proportions of males and females in the full data set and in the sample recommended by the DA tool, the user is less likely to recognize the potential discrimination hidden in the recommendation of the DA tool.

It is noteworthy that such a potentially biased recommendation can be generated due to several reasons such as when the data includes historical records of discrimination against women and the analysis includes variables (e.g., prior annual reviews, prior promotion decisions) in the analysis that bring in the prejudice of previous decision makers in the organization or variables that more adversely impact females than males (e.g., tenure as females are more likely to go on parental leave than males). The above-described scenario falls within Category 2 of Figure 2.1 (i.e. possible discrepancies in the data) discussed earlier, where the full data set includes a balanced representation of demographic classes yet the DA recommendations is unbalanced against a demographic class.

Providing users with aggregated demographic information about the composition of the full data set and the recommended sample put forth by the DA tool will enable users to scrutinize the recommendations of a DA tool in regard to discriminations against a protected class (e.g., females) as it will alert users about the discrepancies between the recommended and/or original data sets. The user can then examine whether there are legitimate reasons for these discrepancies. As was discussed in Chapter 2, the discrepancies in the data analysis data sets can be categorized into four main groups.

Category 1 includes instances where the full data set and the recommended sample are both unbalanced. Using the DT guidance as suggested in this study enables DA users to learn about such discrepancies by examining the aggregated demographic information of all applications in the original full data set and the DA recommended sample. Although the proportions of a protected class in the original data set and the recommended sample might match to some extent, DA users receiving a DT exhibit that includes a disproportionately large share of one demographic class in the original full data set can get alerted to this situation. Category 2 represents instances where the full data set is balanced; yet, the DA-recommended sample is not. In such a situation, DA users by examining the suggested DT guidance can recognize the possible discrimination in the recommended sample and look for the reason(s) for its occurrence. Category 3 represents instances where the full data set is biased; however, the recommended sample is not. Such a situation might occur if a DA tool includes some features (e.g., amended weights of factors) to prevent discrimination. In such cases demographic transparency can help DA users to recognize if a problem exists in the full data set. They can then take steps to rectify the original data set (if possible) to prevent possible discriminatory recommendations based on this set in the future. Finally, under category 4 neither the full data set nor the recommended sample is biased. In such a situation, it is unlikely that discrimination has taken place and the role of DT guidance is to assure users of that.

In light of the above discussion, this study suggests that increasing the level of demographic transparency by providing data analytics users with DT exhibits, as discussed above, increases the user's ability to recognize traces of discrimination (if any) in the system's recommendations. In addition to increasing users' ability, this study suggests that providing demographic transparency, as discussed above, increases the proximity that the DA user feels toward the subjects of their decision. The reason for this increase in the perceived proximity is rooted in the notion of mental construal level, as suggested by construal level theory.

Construal level theory (CLT; Trope and Liberman 2003; Trope and Liberman 2010) suggests that people use a more abstract, high construal level when perceiving, predicting, and judging more psychologically distal targets, and they judge more abstract targets as being more psychologically distal (Bar-Anan et al. 2006). Psychological distance is an egocentric, "subjective experience that something is close or far away from the self, here, and now" (Trope and Liberman 2010, p. 440). Trope and Liberman (2010) discuss the relationship between psychological distance and construal level. They suggest that because less information is available about distal entities than proximal ones, people typically construe the former more abstractly than the latter. In addition, since high-level construals are more general, they tend to bring to mind more distal instantiations of the entities. Therefore, an association is made between the psychological distance and construal level. The greater the psychological distance, the more likely are the perceivers to form high-level rather than low-level construals of objects and vice versa (Trope and Liberman 2010). It is noteworthy, though, that as Trope and Liberman contend psychological distance and construal level are two related yet distinct notions. Psychological distance refers to the perceptions of *when* an event occurs, *where* it occurs, to *whom* it occurs, and *whether* it occurs. However, construal levels refer to the perception of *what* will occur, the representation of the event itself.

In line of the above discussion, this study suggests that demographic transparency increases the perceived proximity that a DA user feels toward the subjects of their decision through decreasing their mental construal level of the population of those subjects (i.e., original pool and recommended sample). This is because high-level construals are viewed as "relatively abstract, coherent, and superordinate mental representations, compared with low-level construals" (Trope and Liberman 2010, p. 441). Therefore, the less abstract and congruent a group of people are perceived to be by an individual, the higher the chance that he/she will have a lower level construal of them. In fact, the number of categories one uses to classify objects and events has been used as an inverse measure of his/her construal level (Henderson et al. 2006; Liberman et al. 2002; Nussbaum et al. 2006; Wakslak et al. 2006). We already know that construal level is tightly related to psychological distance. Hence, when provided with a number of categories that classifies the subjects of analysis to smaller groups, DA users are more likely to feel less psychologically distant from them. Consider, for example, being provided with a number of records that are selected from a larger pool of customers (as the recommended sample to receive a promotion) and receiving no information about the sample in terms of its demographic characteristics. In such a situation, the chance of not thinking through the characteristics of the sample and assuming it as a coherent homogeneous group of people would be higher. However, if further information about the sample's sex, age, marital status, etc. is presented to the user, he/she will be more likely to make a lower level construal of the group of individuals impacted by the analysis and decision. As a result, chances will be higher that he/she will feel less psychologically distant and more proximate to the subjects of the decision. Therefore, this study posits that demographic transparency increases the proximity that a data analytics user feels toward the subjects of his/her analysis/decision.

The fact that in addition to ability, users should be motivated to attend to the provided information is an important pillar of both ELM (Elaboration Likelihood Model) (Petty and Cacioppo 1986a; Petty and Cacioppo 1986b) and cognitive elaboration model of ethical decision-making (Street et al. 2001). Without such motivation, the individual might be reluctant to expend cognitive resources and engage in issue-relevant thinking and as a result, it is less likely that he/she will become aware of the moral aspect of the decision at hand. The moral intensity of an issue can act as a situational motivator (Street et al. 2001). Issues that are more morally intense are more salient and vivid and as a result are easier to notice (Jones 1991). Proximity, an aspect of moral intensity that is of interest in this research, increases saliency and vividness of an issue by bringing to the front the fact that the decision will impact those who are psychologically closer to the decision maker.

This study has already suggested that presenting a data analytic user with demographic transparency is likely to decrease the psychological distance that they feel toward the subjects of their decision. However, due to the importance of proximity in the context of decision making using computer systems in general and data analytics system in particular, this study suggests increasing the level of demographic transparency at the individual level by adding a photo of each data subject next to her/his record. This suggestion is rooted in the differences between texts and pictures as described by Construal Level Theory. Pictures bear physical resemblance to the referent objects, whereas words are abstract representations that carry the essence of that object (Amit et al. 2009; Amit et al. 2008). Hence, words comprise a higher level of construal than do pictures. As high level construals are broad, they tend to bring to mind more distant instantiations of objects. On the other hand, low-level construals are narrow, thus they bring to mind more proximal instantiations of objects. Therefore, pictures as examples of low-level construal are more likely to carry a feeling of proximity to the receiver whereas words as examples of high-level construal are used to represent distal objects in time, space, and social perspectives (Amit et al. 2009).

In addition, pictures are subject to perceptual analysis similar to that performed on the real objects themselves (DeLoache et al. 2003; DeLoache et al. 1998; Stenberg 2006). As perception presupposes proximity of the referent objects, and because pictures are too perceived, they convey a feeling of proximity. The association between visual processing mode and lower level construals as well as decreased psychological distance has been demonstrated in previous research (e.g., Yan et al. 2016). The resemblance of pictures to real objects makes including photos of the human subjects impacted by decisions drawing on data analytics recommendation even more important. This is because such tools represent individuals by a set of rows, columns, and attributes and therefore deprive individuals of their individuality. As a result, the use of data analytics systems has been suggested to be associated with dehumanization (Ebrahimi et al. 2016). Dehumanization, which represents “the denial of qualities associated with meaning, interest, and compassion" toward others (Barnard 2001, p. 98), is a cognitive mechanism facilitating unethical behavior (Bandura 1999) by nullifying self-restraints that operate through feelings of empathy and compassion towards the victims of the unethical behavior in question (Osofsky et al. 2005). Empathy, according to the construal level theory, is a low-level emotion as it does not require construal or transcending one's direct experience (Liberman et al. 2007) and therefore, is more likely to diminish as the perceived psychological distance increases.

It is generally believed that the intensity of affective reactions decreases as the psychological distance increases. For instance, people typically react more strongly to events that are closer to them in terms of time, space, as well as the events that happen to themselves than others. The impact of increases in distance on engaging in unethical behavior has been extensively studied. For instance Bandura (1999, p. 199) argues that "it is easier to harm others when their suffering is not visible and when injurious actions are physically and temporally remote from their effect". In addition, Milgram (1974) in his well-known series of obedience experiments, empirically shows that distance is significantly and inversely related to committing unethical behavior. This study, therefore, suggests that by including photos of data subjects, which leads to decreasing the psychological distance that the user feels toward the subjects of their decision, users will become more likely to realize the moral aspect of the issue at hand.

In light of the above discussion, the following two hypotheses are advanced:

Hypothesis 1: *In the context of a potentially discriminatory data analytics recommendation, DA tools providing higher demographic transparency will increase the likelihood of users' recognition of the moral issue.*

Hypothesis 2: *In the context of a potentially discriminatory data analytics recommendation, DA tools providing higher demographic transparency will increase users' perceptions of proximity toward the subjects of their decision.*

3.3.2. User's Mindfulness

Many unethical decisions stem from a lack of awareness and moral awareness is the critical first step in the ethical decision making process (Rest 1986). Mindfulness is a notion that goes hand-in-hand with awareness and is defined as "keeping one's consciousness alive to the present reality" (Hanh 1976, p. 11). Consciousness encompasses both awareness and attention, where awareness is continuously monitoring the inner and outer environments and attention is the process of focusing conscious awareness and providing increased sensitivity to a small range of experience (Westen 1999). Therefore, mindfulness can be considered as "an enhanced attention to and awareness of current experience or present reality" (Brown and Ryan 2003, p. 822). It is notable that mindfulness is not a stable trait but rather can be increased by training programs and clinical interventions (e.g., Hayes et al. 1999; Kabat-Zinn 1982; Linehan 1993; Segal et al. 2002).

Mindfulness has been shown to have an inverse relationship with willingness to engage in unethical behavior (Ruedy and Schweitzer 2010). This relationship can stem from two reasons. First, several authors have stated that mindfulness is associated with a nonjudgmental, accepting awareness of the environment (e.g., Bishop 2002; Kabat-Zinn 1994). Therefore, mindfulness allows and even encourages taking into account all the relevant information for a given decision. As a result, mindful individuals feel less obliged to ignore or explain away ideas that could be threatening to them (due to a conflict of interest or a potential bias, for example) (Ruedy and Schweitzer 2010). Second, self-awareness is a key facet of mindfulness; therefore, highly mindful individuals are more likely to be aware of and attentive to their internal processes and events and therefore are more likely to engage in effective self-regulation (Ruedy and Schweitzer 2010).

This study suggests that the relationships between demographic transparency and recognition of the moral issue and proximity are positively moderated by users' mindfulness. There are two reasons to this expectation. First, mindfulness captures a quality of consciousness that is mainly thought of as the vividness and clarity of one's current experience and functioning and hence, is in contrast to the mindless, habitual or automatic functioning (Brown and Ryan 2003). Therefore, individuals with higher levels of mindfulness are more likely to scrutinize the information they receive through demographic transparency and recognize the moral aspect of the situation in case the data analytics recommendation includes traces of discrimination. In addition, mindfulness and its focus on attention to the concrete aspects of the present experience (Kabat-Zinn 1994) as well as its emphasis on "being" rather than "achieving" (Kabat-Zinn and Hanh 1990) makes an individual high in mindfulness feel more proximate toward the subjects of their decisions when provided with statistical demographic information about and photos of those individuals.

Second, as Self-determination Theory posits, awareness, in contrast to automatic processing, is a facilitating factor in the choice of behaviors that are consistent with one's interests, values, and needs (Ryan and Deci 2000). As a result, highly mindful individuals are expected to be more inclined than less mindful individuals to discern the discriminatory nature of the data analytics recommendation (if it exists) when provided with the information on the composition of the original pool and the recommended sample in terms of various demographic classes. As a result,

Hypothesis 3: *In the context of a potentially discriminatory data analytics recommendation, users' mindfulness moderates the relationships between demographic transparency and (a) recognition of the moral issue; (b) perceived proximity towards the subjects of their decision, such that the effects are stronger for individuals higher in mindfulness than for those lower in mindfulness.*

3.3.3. Ethical Culture

Ethical culture refers to "a subset of organizational culture, representing a multidimensional interplay among various "formal" and "informal" systems of behavioral control that are capable of promoting either ethical or unethical behavior" (Treviño et al.

1998, p. 451). The positive impact of the organization's ethical culture on its employees ethical decision-making and behavior is well studied in the literature (For a review, see Craft 2013; Loe et al. 2000; O'Fallon and Butterfield 2005). In an early study, Weaver and Ferrell (1977) report the positive impact that existence and enforcement of a corporate policy on ethics can exert on marketing practitioners' ethical beliefs (Loe et al. 2000). Since then, many other business ethics researchers have empirically demonstrated the association between an organization's ethical culture and its employees' moral awareness (e.g., Moberg and Caldwell 2007; Rottig et al. 2011) and in more general terms their moral behavior (e.g., Sweeney et al. 2010; Zhang et al. 2009). Indeed, culture influences individuals' beliefs (Ferrell and Gresham 1985; Hunt and Vitell 1986) and establishes boundaries between what is considered legitimate or unacceptable in an organization (Treviño et al. 1998). Moberg and Caldwell (2007) empirically show individuals' exposure to an organizational ethical culture to be strongly associated with their level of moral imagination (i.e., a process of considering the ethical elements of a decision thoroughly).

These results suggest that when an organization's culture is strong and creates "distinctive, central, and enduring" cognitive images, individuals often adapt their behavior to act in accordance with them (Dutton et al. 1994). From a cognitive perspective, particular thought processes can be invoked by individual's exposure to cultural cues (Gioia and Thomas 1996; Moberg and Caldwell 2007). Previous research has suggested that cultures exert their influence not only by delineating what behavior is right or wrong but also by prompting certain ways of thinking (Moberg and Caldwell

2007). For instance, Hong and her colleagues in a series of studies showed that as a result of exposing bicultural individuals (people who have internalized two cultures) to icons salient to one culture, thought processes associated with the corresponding culture are evoked (Hong et al. 2003; Hong et al. 2000). As a result, this study suggests that in an organization where salient values prompt ethical thinking, it is more likely that ethical thought processes will be invoked in employees' minds as a result of being exposed to demographic transparency. As a result,

Hypothesis 4: *In the context of a potentially discriminatory data analytics recommendation, Ethical culture of the organization moderates the relationship between demographic transparency and recognition of the moral issue, such that the effect is stronger for individuals from organizations with stronger ethical cultures.*

3.3.4. Recognition of the Moral Issue, Proximity, and Approval of the Discriminatory Recommendations

Ethical reasoning has been described as a systematic framework that involves making principled assessment in questionable situations (Ferrell et al. 1989; Rest 1986; Wotruba 1990). Individuals first realize an ethical situation, which prompts them to consider and evaluate courses of actions based on their morality. Such assessments subsequently affect their ethical intentions and actions (Loe and Weeks 2000; Rest 1986; Wotruba 1990). Many studies to date have found significant relationships between the aforementioned stages (For a review, see Craft 2013; Loe et al. 2000; O'Fallon and Butterfield 2005).

Therefore, it is reasonable to expect that users who become morally aware of a potentially discriminatory recommendation of a DA system they are using are more likely not to accept that recommendation compared to users who are not aware. Hence,

Hypothesis 5: *In the context of a potentially discriminatory data analytics recommendation, users' recognition of the moral issue is negatively associated with their acceptance of the system's recommendation.*

Jones (1991) made a significant contribution to the ethical decision making literature by introducing the notion of moral intensity that is about how characteristics of a moral issue itself can impact the ethical decision making process by individuals. Since then, the positive impact of the six aspects of moral intensity (i.e., magnitude of consequences, social consensus, probability of effect, temporal immediacy, proximity, and magnitude of consequences) on various stage of individuals' decision making process (i.e., recognition of the moral issue, moral judgment, moral intention, and moral behavior) as suggested by Rest (1986) have been shown in several studies. This is because moral intensity of an issue influences the saliency and vividness of the ethical issues and as a result affects individuals' attitude and behavior toward them.

Proximity, one dimension of moral intensity, is defined as the "feeling of nearness (social, cultural, psychological, or physical) that the moral agent has for victims (beneficiaries) of the evil (beneficial) act in question" (Jones 1991, p. 376). This suggestion is in line with Hunt and Vitell's (1986) concept of "importance of stakeholders" as an influential variable on the ethical judgment of marketers. Jones posits that people care more about others who are close to them than they do for others who are distant. The positive impact of proximity on ethical recognition as well as ethical behavior has been observed in several studies (e.g., Carlson et al. 2002; Karacaer et al. 2009; Leitsch 2006). Therefore, users who perceive more proximity toward the subjects of a DA analysis are more likely to recognize the moral aspect of a discriminatory recommendation and to consequently not approve it. Thus,

Hypothesis 6: *In the context of a potentially discriminatory data analytics recommendation, users who perceive more proximity toward the subjects of their decision are more likely to recognize the moral aspect of the issue.*

Hypothesis 7: *In the context of a potentially discriminatory data analytics recommendation, users who perceive more proximity toward the subjects of their decision are less likely to accept a DA discriminatory recommendation.*

4. RESEARCH METHODOLOGY

The hypotheses proposed in the present study were tested through a single factor experimental approach by manipulating the level of demographic transparency between participants. Thus, each participant was randomly assigned to one of the three groups: (i) DT-0: no demographic transparency; (ii) DT-1: demographic transparency in the form of charts; and (iii) DT-2: demographic transparency in the form of charts as well as individual photo accompanying each data subject.

4.1. Generating Discriminatory DA Recommendation

A fictitious experimental DA tool was developed that included 200 records of individuals who work in the sales department of an organization. The aim of the analysis was to generate a list of 20 individuals to be sent to a training program on effective leadership in a sales organization. To generate the list, the system drew on various objective (education level and years of working experience at the company), and subjective factors (average of performance evaluation over the last 3 years and potential of the employee). Participants were told that the subjective factors had been provided by employees' managers. More specifically, evaluation of the performance and potential of the employee had been provided by the employee's supervisor at the end of every year. Figures 4.1 and 4.2 respectively depict snapshots of the pages that provided participants with definitions of the variables in the system and a list of modeling parameters that would be taken into account to generate the recommendations.

The recommended sample of employees to be sent to the training program included discrimination against women as the proportion of female individuals in the recommended sample was considerably reduced compared to its level in the full data set (44% in the full data set versus 15% in the recommended sample). The discriminatory recommendation was generated following the literature that suggest when labeling the data (e.g., defining a good employee), if one or several of the defining variables are subjective, they might bring in the personal prejudice of previous managers into the analysis process (Barocas and Selbst 2016; Calders and Žliobaitė 2013; Stauffer and Buckley 2005). In the experiment, the prejudice of the manager toward females had led to their receiving lower performance evaluations as well as evaluation of potential of the employee compared to their male counterparts. Since the recommendation of the tool took into account these two variables and was indeed generated based on a linear function of the four aforementioned variables (See Figure 4.2), the recommendation included discrimination against females. It is noteworthy that this is an example of indirect discrimination (a.k.a., disparate impact) as the discriminatory recommendation is generated based on variables that are neutral on the surface. In other words, the analysis does not take into account employees' sex but due to the hidden relationships in the data, generates an outcome that includes discrimination against females.

Variables' Descriptions

Below are the variables in the data set you just received along with a short description of each.

Variable	Description
ID	The employee's ID in the database.
Current Department	The name of the department at which the employee is currently employed.
Name	The employee's name.
Level of Education	The level of highest education the employee has earned (e.g., Bachelors, Masters, and Ph.D.)
Years of working experience at the current company	Number of years the employee has worked in this organization
Average of performance evaluation over the last 3 years	Average of the employee's performance evaluation in the last three years. Performance evaluation is provided by the employee's supervisor at the end of every year. 1 represents "Unsatisfactory" 2 represents "Below Expectations" 3 represents "Meets Expectations" 4 represents "Exceeds Expectations" 5 represents "Outstanding"
Potential of the employee	The appraisal of the employee's potential for learning and new roles and responsibilities. This evaluation was provided by the employee's supervisor in the performance evaluation last year. 1 represents "Unsatisfactory" 2 represents "Below Expectations" 3 represents "Meets Expectations" 4 represents "Exceeds Expectations" 5 represents "Outstanding"

Figure 4.1. Description of the Variables in the Experimental DA Tool

Modelling Parameters

Below is the list of variables that, according to historical data, are most associated with success in a training similar to the one in question. These variables are:

- **Level of Education**
- **Years of Working Experience at the Current Company**
- **Average of performance evaluation over the last 3 years**
- **Potential of the employee**

Figure 4.2. Modeling Parameters in the Experimental DA Tool

4.2. Design of Demographic Transparency Decisional Guidance

For participants in the two groups with demographic transparency (i.e., DT-1 and DT-2), a step was added prior to accepting (or rejecting) the data analytics recommendation. The step included demographic transparency decisional guidance, which aims at increasing the participants' ability and motivation to scrutinize the information further in terms of ethical considerations. The "automatic" method was used to provide demographic transparency (Gregor and Benbasat 1999). Therefore, in the second and third groups (i.e., DT-1 and DT-2), after providing the user with the recommendations, the pertinent demographic transparency was shown to the user such that the user could not submit their decision (of approving or rejecting the system's recommendation) prior to receiving the demographic transparency screen.

Demographic transparency guidance depicted the proportion of members of each demographic class in both the original pool and the recommended sample by the data analytics tool. This method is in line with discrimination measures discussed in Chapter 2. Specifically, extended lift, a well-used measure for quantifying discrimination which calculates the ratio between the proportion of individuals from a protected group who have received a benefit (or have been denied a benefit) and the proportion of individuals from that group in the data set (Pedreschi et al. 2009; Pedreshi et al. 2008; Romei and Ruggieri 2014; Ruggieri et al. 2010). Therefore, extended lift is related to and depicts how a disadvantaged group is over-represented (under-represented) in a data set of individuals who are granted (denied) a benefit (Pedreschi et al. 2013).

In line with the notion of the extended lift measure, and in order to provide DA users with a decisional guidance that illustrates an unbalanced data set, which includes an over/under-representation of a group of individuals, the demographic transparency chart used in this study was designed such that it depicts the proportion of each sex in both the original dataset and the recommended sample. As can be seen in Figure 4.3, the original data set in this study included 44% females and 56% males. However, the recommended sample of selected employees to be sent to the training program comprised 15% females and 85% males. The demographic transparency chart (exhibited in Figure 4.3) shows the unbalanced recommended sample.

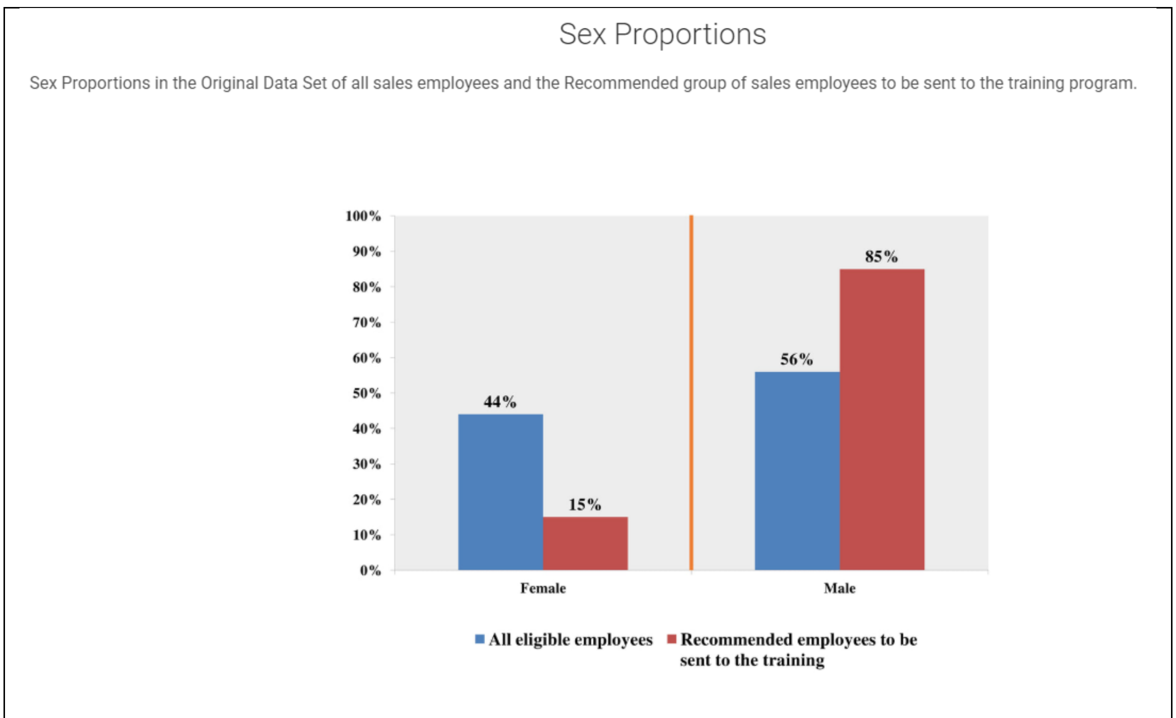


Figure 4.3. Aggregate Level Demographic Transparency Chart (Provided to Participants in Treatments DT-1 and DT-2)

For participants in the DT-2 treatment condition, in addition to the aforementioned aggregated demographic transparency guidance, individual level demographic information was also provided such that a photo of each data subject was presented next to her/his record (See Figure 4.5). This approach is in line with the objective of increasing the level of proximity that DA users feel toward the subjects of the decision.

It is noteworthy that all names and photos of individuals were selected from one race (i.e., white) as including names and photos of individuals from various races could have introduced the confounding impact of race and the possibility of racial discrimination to the study. In addition, it is worth mentioning that although the system's recommendations in this study included potential discrimination against females, participants in the DT-1 and DT-2 conditions received aggregated-level DT exhibits (i.e., charts) about both sex and age information. However, traces of potential discrimination could only be found in the diagram related to sex of the employees. The reason for this approach was to make the real purpose of the study less clear to participants.

The validity of the experimental manipulation for the level of demographic transparency was tested by asking participants to indicate their level of agreement with three statement such as "I felt that I had demographic transparency regarding the subjects of my decision". Appendix A includes the details of the questions used for manipulation check.

4.3. Measures

To ensure content validity, all measurement instruments were adapted from the existing and validated scales in the literature and operationalized using 7-point Likert scales, with the endpoints labeled as "strongly disagree" and "strongly agree". Recognition of the moral issue was measured using Reynolds' (2006) three-item moral awareness scale (e.g., "There are very important ethical aspects to this situation"). Perceived proximity to the subjects of the decisions was measured with Barnett's (2001) three-item scale (e.g., "Compared to yourself, do you believe those potentially affected by the depicted action are dissimilar/similar"). Mindfulness was assessed using Brown and Ryan's (2003) 15-item scale (e.g., "It seems I am running on automatic without much awareness of what I'm doing."). Ethical culture was assessed by adapting a 14-item scale developed by Treviño et al. (1998) (e.g., "Ethical behavior is rewarded in our organization"). Appendix A includes the details of the measurement instruments.

This study controlled for the effect of gender and age of the respondents as some research (e.g., Cohen et al. 2001) has shown that women are more likely to act morally in ethical situations. Ethical behavior has been observed to increase with age as well (e.g., Singhapakdi et al. 1996a). Furthermore, in order to rule out the confounding impact of the level to which one believes in upholding ethical values on the results, this study measured and controlled for participants' moral identity. Moral identity is defined as "a self-consistent commitment to lines of action benefitting others" (Hart et al. 1998, p. 513) and as "a-self-conception organized around a set of moral traits" (Aquino and Reed 2002,

p. 1424). To that end, participants completed the self-importance of moral identity (SMI) scale (Aquino and Reed 2002), which asks participants to consider a person with the following characteristics: caring, compassionate, fair, friendly, generous, helpful, hardworking, honest, and kind. Then, participants are asked to respond to a number of items indicating how important it is to them to be someone who has these characteristics (internalization subscale) and how important it is to them to appear as someone who has these characteristics (e.g., by buying products or reading books that demonstrate these attributes to others) (symbolization subscale).

In addition, since prior knowledge of the possibility of data analytics tools generating discriminatory recommendations might significantly impact moral awareness and behavior, participants' knowledge of the issue was measured (with a single item: "I know quite a bit about discriminatory recommendations of data analytics systems") and controlled for[7].

Last but certainly not least, this study strives to minimize the potential social desirability effect as it has been discussed as an important variable in organizational ethics studies because of their reliance on self-report instruments and its sensitive nature (Randall and Fernandes 1991). Social desirability is mainly manifested in two terms: self-deception, and impression management (Paulhus 1984). Self-deception is defined as the propensity of individuals to "deny having psychologically threatening thoughts or feelings" (Paulhus 1991, p. 4) and impression management is defined as the propensity of respondents to "consciously over-report their performance of a wide variety of desirable behaviors and

[7] To prevent the potential priming effect, this question was asked at the end of the survey.

under-report undesirable behaviors" (Paulhus 1991, p. 4). In order to minimize the potential social desirability effect, participants were assured that the survey is anonymous and confidential. Furthermore, Paulhus' (1991) suggestion to use the Impression Management subscale of the Balanced Inventory of Desirable Responding (BIDR) was followed to test for social desirability effects. The choice of impression management was made to follow existing studies in the business ethics literature (e.g., Flannery and May 2000; Treviño et al. 1998; Watley and May 2004).

4.4. Participants and Sample Size

The sample for this study consists of middle managers who had at least two employees reporting to them. This sampling choice was made since the context of this study is selecting a group of employees to be sent to a training program and therefore, a knowledge of managing and overseeing the works of a few employees seemed appropriate. An invitation to participate in the study was broadcasted by a market research firm via email. Individuals receive a point-based incentive (redeemable for various prizes) for their assistance in the study.

In order to determine the number of required participants, a power analysis was performed, which determined that 153 subjects (51 subjects for each group) would assure a sufficient statistical power of 0.80 to detect a medium effect size (f= 0.25) (Cohen 1988). This number would also satisfy the PLS requirement as PLS demands the sample size to be ten times the number of items used to measure the construct with the highest number of items (Gefen et al. 2000). Since the measurement used for mindfulness included 15 items, the minimum number of participants in regard to PLS analysis was 150.

4.5. Pilot Study

Prior to data collection, a pilot study was conducted to (i) confirm that the demographic transparency messages were perceived as intended; (ii) examine the clarity of the instructions and measurement instruments; and (iii) ensure the technical feasibility of the designed tool and online questionnaires. To that end, the pilot experiment was conducted with 40 middle managers, who were recruited through a market research firm. Participants were asked to perform the online experiment and respond to manipulation check items as well the items used to measure other constructs in the study (e.g., recognition of the moral issue and perceived proximity). In addition, participants were asked to comment on the demographic transparency diagrams and their level of clarity. In addition, any technical difficulty they might have experienced were solicited. The feedback obtained through the pilot study was used to refine the experimental tool and demographic transparency diagrams and photos. The pilot study did not result in any changes in the measurement instrument. Then the following procedures were used to administer the main controlled online experiment for data collection. It is noteworthy that ethics approval was secured from the McMaster Research Ethics Board prior to any data collection.

4.6. Experimental Procedures

Participants for this study were told that an organization wanted to send 20 of its employees abroad to a one-week training program on Effective Leadership in a Sales Organization. To help participants realize the importance of the training, they were told that "attending the training will bring about great experience for the selected employees. In addition, they will be more likely to receive promotions in the future". The participants' task was to use the fictitious data analytics tool and evaluate its recommendation of what employees to be sent to the training program. The use of a fictitious data analytics tool and an online experiment was to control for the experimental setting and to minimize the impact of other variables that could possibly contaminate the results.

Participants progressed through the experiment as follows.

1. Participants were first asked to read the consent letter (Appendix B) and agree to participate in the study.

2. Next, participants read a document that included instructions on how to use their assigned data analytics tool and perform the task (Appendix C).

3. Subsequently, participants were presented with a dataset of the 200 sales employees. It is noteworthy that this list of individuals was different for participants in DT-2 condition from participants in DT-0 and DT-1 conditions. As previously discussed, unlike participants in DT-0 and DT-1, participants in

condition DT-2 received a photo of each employee next to their information. Figures 4.4 and 4.5 depict a snapshot of a portion of the data set that participants received. Names of the fictitious employees were added to the dataset to make the situation similar to a real experience.

4. In the next step, participants were provided with a description about each of the variables that were included in the data set (Figure 4.1).

   Subsequently, participants received the list of the variables that in the past have been most closely associated with success in a training similar to the one in question. These variables were level of education, years of working experience at the current company, average of performance evaluation over the last three years, and potential of the employee (see Figure 4.2).

Show All entries

Search:

ID	Current Department	Name	Education	Years of working experience at the current company	Average of performance evaluation over the last 3 years	Potential of the employee
28	Sales	Audrey Watson	Masters	15	3	3
230	Sales	Peter Walters	Masters	13	4	3
431	Sales	Luke Clark	Bachelors	7	4	5
635	Sales	Heather Carlson	Masters	2	4	2
597	Sales	Megan Berry	High School	4	3	2
427	Sales	Lindsay Cox	Bachelors	7	3	3
54	Sales	Dave Arnold	Masters	14	5	4
49	Sales	Allie Schultz	Masters	14	4	2
643	Sales	Maddy Martin	Masters	1	3	3
489	Sales	Alec Reed	High School	5	4	5

Figure 4.4. A portion of the Data Set of all Employees in the Systems (for Participants in Treatments DT-0 and DT-1)

Show All entries Search:

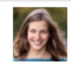
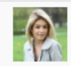
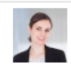
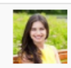
ID	Current Department	Name	Education	Years of working experience at the current company	Average of performance evaluation over the last 3 years	Potential of the employee	
28	Sales	Audrey Watson	Masters	15	3	3	
230	Sales	Peter Walters	Masters	13	4	3	
431	Sales	Luke Clark	Bachelors	7	4	5	
635	Sales	Heather Carlson	Masters	2	4	2	
597	Sales	Megan Berry	High School	4	3	2	
427	Sales	Lindsay Cox	Bachelors	7	3	3	
54	Sales	Dave Arnold	Masters	14	5	4	
49	Sales	Allie Schultz	Masters	14	4	2	
643	Sales	Maddy Martin	Masters	1	3	3	
489	Sales	Alec Reed	High School	5	4	5	

Figure 4.5. A portion of the Data Set of all Employees in the Systems (for Participants in Treatment DT-2)

5. Next, participants clicked on a button that based on those variables, ran a pre-designed predictive model and generated recommendations. The recommendation that participants received was the same list of all employees with one additional attributes for each record: the predicted interest value (0 or 1), which, best upon its prediction of the employee in question to be successful, indicated whether or not they were selected to be sent on the training program. Participants had as much time as they needed to scrutinize the results (e.g., sort and/or filter the records based on various attributes).

6. In the next step, participants in the control condition (DT-0) were asked to indicate their decision on whether they accept the recommendation of the system. However, participants in the two DT conditions, prior to making their decision, received the DT charts (Figure 4.3).
7. Upon completion of the experimental task, participants filled out a questionnaire that includes the measures of the dependent and control variables as well as the manipulation check.
8. Finally, participants were debriefed (Appendix D) and thanked for their participation.

4.7. Model Validation

Structural Equation Modeling (SEM) was used to validate the proposed research model. SEM combines a measurement model (i.e., confirmatory factor analysis) and a structural model (i.e., relationships between constructs of interest) (Meyers et al. 2006). More specifically, Partial Least Squares (PLS) was used as the SEM method to analyze the data and validate the proposed model for two main reasons: (1) PLS, as a component-based SEM technique, is more suited for studies with an exploratory nature (Gefen et al. 2000), which is the case with this study; (2) PLS does not make distributional assumptions regarding the data (Chin 1998; Chin et al. 2003; Venkatesh and Agarwal 2006). To that end, SamrtPLS 3.0 was deployed in this study for purposes of data analysis and model validation. The evaluation of the research model followed a two-step process: (1) the measurement model, and (2) the structural model. Chapter 4 presents the results of both stages in detail. Next, a summary of the analyses performed will be briefly discussed.

To evaluate the measurement model, the focus is placed on the reliability and the validity of the measures used to represent the model's constructs (Chin 2010). Table 4.1 illustrates the tests performed to evaluate the constructs to that end.

Table 4.1. Summary of Test- Measurement Model

Analysis	Test	Note
Reliability of Measurement Instruments	Cronbach's alpha	Acceptance criterion: Value > 0.70 (Nunnally and Bernstein 1994)
	Composite reliability	Acceptance criterion: Value > 0.60 (Bagozzi and Yi 1988)
Convergent and Discriminant Validity	Item cross-loading	Acceptance criterion: The loading on corresponding construct (i.e., theoretical construct) should be larger than loading on other constructs by at least 0.10 (Chin 2010; Gefen and Straub 2005)
	Fornell-Larcker Criterion	Acceptance criterion: The square root of the Average Variance Extracted (AVE) of a construct must be larger than the correlation between that construct and any other construct in the model (Barclay et al. 1995)
Multicollinearity	- Bivariate Correlations - VIF	Acceptance criteria: - Bivariate correlations greater than 0.8 can indicate traces of multicollinearity (Meyers et al. 2006) - Variance Inflation Factors (VIFs) greater than 3.3 may indicate potential multicollinearity issues (Petter et al. 2007)

After the appropriateness of the measurement model was established, the structural model was evaluated to determine whether the proposed research model is supported by the data collected (Chin 2010). Table 4.2 provides a summary of the analyses performed to evaluate the structural model.

Table 4.2. Summary of Test- Structural Model

Test	Calculation	Note
Path Coefficients Significance	Obtained from SmartPLS	Bootstrap approach was employed to evaluate the significance of path coefficients (Chin 1998)
R^2 for endogenous Variables	Obtained from SmartPLS	Although no specific acceptable threshold value has been set for R^2, a large enough R^2 values to achieve Adequate explanatory power is sought-after (Gefen et al. 2000; Urbach and Ahlemann 2010)
Effect Sizes	Obtained from SmartPLS	The magnitude of the effect sizes of each path was evaluated following these values: f^2_{small} (.02), f^2_{medium} (.15), and f^2_{large} (.35) (Chin 2010)
Goodness of Fit (GoF) index	$GOF = \sqrt{\overline{Communality} \times \overline{R^2}}$	Absolute GoF can be used to assess the PLS model in terms of overall (both measurement and structural levels) prediction performance The suggested baseline values of GoF_{small} (.10), GoF_{medium} (.25), and GoF_{large} (.36) were used to evaluate fit of the model (Tenenhaus et al. 2005; Wetzels et al. 2009)

Subsequent to the above analysis to evaluate the structural model, the following post-hoc analyses were also conducted:

1. ANOVA analysis to compare the differences between the three treatment groups (DT-0, DT-1, and DT-2) in terms of the endogenous variables in the study.

2. ANOVA analyses to shed light on the moderating effects of user's mindfulness and ethical culture.

3. The effects of control variables that were captured in the study (i.e. respondents' gender, age, moral identity, impression management, and prior knowledge of the possibility of data analytics tools generating discriminatory recommendations) were examined.

The next chapter will present in detail the data analyses performed in this dissertation, as well as the results obtained.

5. RESULTS

To test the proposed hypotheses, two main methodologies were employed: Structural Equation Modeling (SEM) and analysis of variance (ANOVA). To that end, first, a series of preliminary data analyses were performed to examine the quality of the measurements used. Next, to test the seven stated hypotheses, SEM analyses were performed using SmartPLS 3.0. Subsequently, ANOVA tests were undertaken to examine the impacts of various levels of DT on the two mediating variables. In addition, ANOVA tests were conducted to shed more light on the moderating impacts of user's mindfulness and ethical culture.

5.1. Preliminary Data Analyses

In order to ensure that statistical and methodological artefacts do not impact the measurements, analyses, and results, a series of preliminary data analyses were conducted. According to the existing literature, there are five main factors that impact the quality of measurements, experimental manipulations, and analyses in this study: (i) presence of outliers (e.g., Barnett and Lewis 1974), (ii) low reliability of variables measurements (e.g., Nunnally 1978), (iii) low validity of factors (e.g., Straub et al. 2004) (iv) multicollinearity among the factors (e.g., Meyers et al. 2006), and (v) common method bias (e.g., Straub et al. 2004).

5.1.1. Outliers Analyses

Outliers are "cases with extreme or unusual values on a single variable (univariate) or on a combination of variables (multivariate)" (Meyers et al. 2006, p. 65). Analyses were performed to detect both types of outliers: univariate and multivariate.

To detect univariate outliers, composite scores were calculated for each of the constructs and, employing IBM SPSS Statistics 24, box plots were prepared. Overall, a total of 7 unique cases with univariate outliers were identified. Interestingly one case (ID 110) appeared as outliers in two constructs: proximity and ethical culture and other cases appeared as outliers only in one construct: ethical culture. In order to deal with these 7 cases, each case was scrutinized in detail. All respondent seemed to have spent enough time on responding to the questionnaire. In addition, the pattern of responses in all cases other than case 110 did not show any sign of frivolity. Therefore, Meyer et al.'s (2006, p. 66) suggestion about asking the fundamental question of "Does this outlier represent my sample?" was followed. As a result, all cases of univariate outliers other than case 110 were retained in the data set as it is possible that a number of respondents came from organizations, which have lower ethical culture compared to the average of the respondents.

A Mahalanobis distance analysis was also conducted to detect multivariate outliers. This metric measures "the multivariate "distance" between each case and the group multivariate mean (known as centroid)" (Meyers et al. 2006, p. 67). Mahalanobis distance was calculated for each case and compared with the chi-square distribution (alpha level =

0.001). A case can be considered a multivariate outlier if its Mahalanobis distance is above this threshold. After conducting this analysis, only one multivariate outlier was detected (case ID: 110). Therefore, this case was removed from the data set as it was found to be a univariate and multivariate outlier.

5.1.2. Reliability Analysis

To ensure about the consistency of the measurements, the reliabilities of the factors were tested. Reliability refers to the extent to which a set of measurement items is consistent in measuring the pertinent factor (e.g., recognition of the moral issue) (Pedhazur and Schmelkin 1991; Straub et al. 2004). To that end, SPSS 24 was employed to calculate Cronbach's alpha (Cronbach 1951) for each of the factors (see Table 5.1). In addition, composite reliability (CR) as calculated by SmartPLS 3.0, is also reported for each variable in Table 5.1. All factors exhibited acceptable reliability ($\alpha \geq 0.7$) (Kline 2000; Nunnally 1978) and ($CR \geq 0.6$) (Bagozzi and Yi 1988); as such, it can be concluded that the data analysis results will not suffer from measurement error introduced by low reliabilities of measures.

Table 5.1. Reliability Statistics		
Factor	**Internal Consistency (Cronbach's Alpha)**	**Internal Consistency (Composite Reliability)**
Recognition of the moral issue	0.94	0.96
Perceived proximity	0.82	0.89
User's mindfulness	0.92	0.94
Ethical culture	0.95	0.95

5.1.3. Validity Analyses

Construct validity of a factor indicates whether its measurement items adequately correlated with one another (i.e., convergent validity) and discriminate the factor from other factors in the study (i.e., discriminant validity) (Pedhazur and Schmelkin 1991; Straub et al. 2004). Two techniques of cross-loading analysis and Fornell-Larcker criterion are commonly used to assess construct validity (Fornell and Larcker 1981; Urbach and Ahlemann 2010).

For the purpose of the first technique, cross-loading analysis, each factor's component scores are correlated with all the measurement items. If each item loading is larger for its pertinent factor by at least 0.1 and each of the factors loads highest with its pertinent items, it can be inferred that there is an adequate level of construct validity (Chin 1998; Gefen and Straub 2005; Urbach and Ahlemann 2010). SmartPLS 3.0 was used to calculate the loadings between the four factors in the research model of this study (i.e., recognition of the moral issue, perceived proximity, user's mindfulness, and ethical culture) and their measurement items. As depicted in Table 5.2, all of the items had larger loadings on their pertinent factors compared to their loadings on other factors by the difference magnitude of at least 0.1. In addition, all four factors loaded higher with their pertinent items than other items. Therefore, as per the recommendations in the literature, the results have adequate construct validity (Hair et al. 2010; Meyers et al. 2006; Straub et al. 2004).

Table 5.2. Items Loadings and Cross-loadings of Measures

Construct	Item	Recognition of the moral issue	Perceived proximity	User's mindfulness	Ethical culture
Recognition of the moral issue	Recog1	**0.937**	0.425	0.256	0.323
	Recog2	**0.964**	0.432	0.263	0.340
	Recog3	**0.931**	0.421	0.249	0.324
Perceived proximity	Prox1	0.433	**0.900**	0.195	0.103
	Prox2	0.350	**0.838**	0.274	0.115
	Prox3R	0.378	**0.842**	0.287	-0.006
User's mindfulness	Mindfulness1R	0.251	0.320	**0.644**	0.148
	Mindfulness2R	0.171	0.251	**0.620**	0.088
	Mindfulness3R	0.116	0.196	**0.716**	0.151
	Mindfulness4R	0.182	0.157	**0.713**	0.053
	Mindfulness5R	0.146	0.121	**0.589**	0.015
	Mindfulness6R	0.191	0.261	**0.648**	0.088
	Mindfulness7R	0.278	0.270	**0.800**	0.109
	Mindfulness8R	0.145	0.191	**0.815**	0.097
	Mindfulness9R	0.257	0.145	**0.723**	0.099
	Mindfulness10R	0.125	0.190	**0.749**	0.031
	Mindfulness11R	0.124	0.163	**0.646**	0.026
	Mindfulness12R	0.108	0.092	**0.635**	0.125

Table 5.2. Items Loadings and Cross-loadings of Measures

Construct	Item	Recognition of the moral issue	Perceived proximity	User's mindfulness	Ethical culture
User's mindfulness (Cont.)	Mindfulness13R	0.240	0.180	**0.757**	0.192
	Mindfulness14R	0.196	0.180	**0.821**	0.104
	Mindfulness15R	0.160	0.160	**0.623**	0.087
Ethical culture	Culture1	0.244	0.098	0.135	**0.688**
	Culture2R	0.169	-0.080	0.040	**0.689**
	Culture3	0.228	0.085	0.108	**0.806**
	Culture4	0.217	0.071	0.069	**0.812**
	Culture5	0.297	0.026	0.065	**0.852**
	Culture6	0.396	0.194	0.214	**0.828**
	Culture7R	0.243	0.075	0.110	**0.741**
	Culture8	0.338	0.104	0.147	**0.857**
	Culture9R	0.188	-0.022	0.059	**0.674**
	Culture10	0.232	0.015	0.102	**0.782**
	Culture11	0.333	0.022	0.063	**0.878**
	Culture12R	0.104	-0.069	0.029	**0.659**
	Culture13	0.342	0.131	0.174	**0.846**
	Culture14	0.162	-0.059	0.059	**0.620**

For the second technique, the Fornell-Larcker criterion (Fornell and Larcker 1981) requires a factor to share more variance with its assigned measurement items than with any other factor. Accordingly, the Average Variance Extracted (AVE) of each factor should be larger than the factor's largest correlation with any other factor (Gefen and Straub 2005; Lehmann 1988; Urbach and Ahlemann 2010). In this study, SmartPLS 3.0 was employed to compute the AVEs and correlations for the four factors. The results, depicted in Table 5.3, illustrate adequate construct validity for all four factors.

Table 5.3. Factors' Correlations and Square Roots of AVE for Discriminant Validity				
	Recognition of the moral issue	**Perceived proximity**	**User's mindfulness**	**Ethical culture**
Recognition of the moral issue	0.944			
Perceived proximity	0.451	0.861		
User's mindfulness	0.272	0.292	0.704	
Ethical culture	0.348	0.079	0.141	0.771

5.1.4. Multicollinearity Analysis

Multicollinearity is the extent to which a factor can be explained through other factors in the analysis. High multicollinearity among factors in an analysis causes an underestimation of the effect of any single factor, owing to their strong interrelationships (Hair et al. 2010). To assess multicollinearity among the four factors used in the research model (i.e., recognition of the moral issue, perceived proximity, user's mindfulness, and ethical culture), an examination of the inter-factor correlations and Variance Inflation Factor (VIF) was made (Meyers et al. 2006). All of the four factors in the model exhibited bivariate correlations of less than 0.5 and VIF values less than 1.5 (Meyers et al. 2006). Hence, it can be concluded that multicollinearity was not an issue in the analysis of this study (Diamantopoulos and Siguaw 2006; Tabachnick and Fidell 1996).

5.1.5. Common Method Bias Analyses

Common Method Bias (CMB) refers to the potential variance in the self-report factors that are attributable to the measurement method and not the hypothesized relationships between the factors (Podsakoff et al. 2003; Straub et al. 2004). Such a bias is a potential threat to the validity of the findings in an empirical study (Burton-Jones 2009; Podsakoff and Organ 1986). To assess the level of common method bias in this study, two techniques were drawn upon: (i) "Herman's One-factor Test", and (ii) unmeasured latent method construct (ULMC)[8].

To conduct Harman's one-factor test, all the measurement items of the four factors were entered into an exploratory principal components analysis (PCA). Then, to assess the existence of common method bias, the results were scrutinized. Common method bias exists if (1) all of the items tend to load on one single general factor, or (2) one factor explains more than 50% of the variance in all of the items (Podsakoff and Organ 1986). Results of Harman's one-factor test did not show the presence of common method bias in this study as the unrotated solution to PCA for the items suggested five factors and the largest variance explained by one factor was 27.2%.

Moreover, the presence of common method bias can be assessed using ULMC method (Podsakoff et al. 2003; Williams et al. 2003). The method, delineated by Liang et al. (2007) in detail, was followed in this study. In this method, three steps were followed: (1)

[8] Although I acknowledge the critique of Chin et al. (2012) of the unmeasured latent method construct approach suggested by Liang et al. (2007), following other recent studies in the literature, the test was performed in this study (e.g., Chen and Shen 2015; Maier et al. 2015; Wang et al. 2016) because it might detect CMB influence (Maier et al. 2015)

each measurement item was used to create a single-item construct, (2) each original construct in the model (e.g., Recognition of the moral issue) was linked to its pertinent single-item constructs (e.g., Recog1), and (3) a method construct with all the items was added to the research model, by linking it to each single-item construct. Subsequently, the model was run using SmartPLS 3.0 and the coefficients of the paths from the substantive construct (i.e., theoretical construct) and the method factor to each single-indicator construct (denoted as R_1 and R_2, respectively) were examined (see Table 5.4). Since "the squared values of the method factor loadings were interpreted as the percent of indicator variance caused by method, whereas the squared loadings of substantive constructs were interpreted as the percent of indicator variance caused by substantive constructs. If the method factor loadings are insignificant and the indicators' substantive variances are substantially greater than their method variances, it can be concluded that common method bias is unlikely to be a serious concern" (Liang et al. 2007, p. 87). As a result, no traces of common method bias was found in the study using the ULMC method as (1) no items had a significant method factor loading (at $p < 0.05$), while the entire substantive construct loadings were significant ($p<0.001$), (2) the average substantive variances (0.6) was considerably larger than the average method variances (lower than 0.01).

Table 5.4. ULMC Common Method Bias							
Construct	**Indicator**	**R_1**	**Sig.**	**R_1^2**	**R_2**	**Sig.**	**R_2^2**
Recognition of the moral issue	Recog1	0.939	p<0.001	0.882	0.001	n.s.	0.000
	Recog2	0.963	p<0.001	0.927	0.002	n.s.	0.000
	Recog3	0.931	p<0.001	0.867	-0.003	n.s.	0.000
Perceived proximity	Prox1	0.907	p<0.001	0.823	-0.011	n.s.	0.000
	Prox2	0.838	p<0.001	0.702	0.048	n.s.	0.002
	Prox3R	0.837	p<0.001	0.701	-0.037	n.s.	0.001
User's mindfulness	Mindfulness1R	0.510	p<0.001	0.260	0.133	n.s.	0.018
	Mindfulness2R	0.575	p<0.001	0.331	0.031	n.s.	0.001
	Mindfulness3R	0.685	p<0.001	0.469	0.053	n.s.	0.003
	Mindfulness4R	0.747	p<0.001	0.558	-0.057	n.s.	0.003
	Mindfulness5R	0.656	p<0.001	0.430	-0.095	n.s.	0.009
	Mindfulness6R	0.618	p<0.001	0.382	0.018	n.s.	0.000
	Mindfulness7R	0.768	p<0.001	0.590	0.032	n.s.	0.001
	Mindfulness8R	0.849	p<0.001	0.721	-0.031	n.s.	0.001
	Mindfulness9R	0.730	p<0.001	0.533	0.012	n.s.	0.000
	Mindfulness10R	0.827	p<0.001	0.684	-0.087	n.s.	0.008
	Mindfulness11R	0.713	p<0.001	0.508	-0.086	n.s.	0.007
	Mindfulness12R	0.667	p<0.001	0.445	0.017	n.s.	0.000
	Mindfulness13R	0.668	p<0.001	0.446	0.123	p<0.1	0.015
	Mindfulness14R	0.852	p<0.001	0.726	-0.026	n.s.	0.001
	Mindfulness15R	0.645	p<0.001	0.416	0.001	n.s.	0.000

Table 5.4. ULMC Common Method Bias							
Construct	**Indicator**	**R_1**	**Sig.**	**R_1^2**	**R_2**	**Sig.**	**R_2^2**
Ethical culture	Culture1	0.609	p<0.001	0.371	0.101	n.s.	0.010
	Culture2R	0.815	p<0.001	0.664	-0.132	n.s.	0.017
	Culture3	0.830	p<0.001	0.689	-0.015	n.s.	0.000
	Culture4	0.889	p<0.001	0.790	-0.080	n.s.	0.006
	Culture5	0.899	p<0.001	0.808	-0.063	n.s.	0.004
	Culture6	0.576	p<0.001	0.332	0.268	p<0.1	0.072
	Culture7R	0.743	p<0.001	0.552	0.012	n.s.	0.000
	Culture8	0.758	p<0.001	0.575	0.106	n.s.	0.011
	Culture9R	0.752	p<0.001	0.566	-0.082	n.s.	0.007
	Culture10	0.798	p<0.001	0.637	-0.019	n.s.	0.000
	Culture11	0.915	p<0.001	0.837	-0.054	n.s.	0.003
	Culture12R	0.840	p<0.001	0.706	-0.181	n.s.	0.033
	Culture13	0.702	p<0.001	0.493	0.156	n.s.	0.024
	Culture14	0.694	p<0.001	0.482	-0.077	n.s.	0.006
Average		**0.764**		**0.597**	**-0.001**		**0.008**

5.2. Data Analysis Results

As discussed in Chapter 4, to test the seven proposed hypotheses, two main methods were employed: Structural Equation Modeling using SmartPLS 3.0 and Analysis of Variance (ANOVA). This section illustrates the results of these analyses, prior to which information about two other topics will be discussed and presented: demographics and backgrounds of the participants, and results of the manipulation check.

5.2.1. Subjects' Background Information

An invitation to participate in the study was broadcasted by a market research firm. The sample for this study consists of 183 (61 for each group) middle managers who had more than one employee reporting to them. According to a power analysis, this number would assure a sufficient statistical power of 0.80 to detect a medium effect size ($f = .25$) (Cohen 1988). In addition, this number fulfills the requirement about minimum number of subjects in a PLS analysis (i.e., ten times the number of items used to measure the construct with the highest number of items). (Gefen et al. 2000).

The 183 subjects were recruited from various industries including but not limited to education, government, healthcare, real estate, information services and data processing, construction, and finance. The subjects were employed in various departments such as human resources, research and development, accounting, marketing and sales, customer service, and IT. Among the subjects 25% were from organization with less than 100 employees, 21% from organization with 100-500 employees, 11% from organization with

500-1000 employees, 16% from organizations with 1000-5000 employees, and 27% from organizations with more than 5000 employees. 90 (49.2 percent) of the subjects were female and 93 (50.8 percent) of the subjects were male. The average age of the participants was 49.5. There was no significant difference in industry, department, organization size, gender, and age distribution across the three treatments.

In general, the subjects were familiar with data analytics tools (mean: 4.1/7). However, their prior knowledge of the possibility of such tools generating discriminatory recommendations were below average (mean: 3.4/7). In addition, the subjects were familiar with equity regulations (mean: 4.6/7). No significant differences were found across the three treatment conditions regarding these three aspects.

5.2.2. Manipulation Check

The independent variable in this study is the level of demographic transparency, which was manipulated to have three levels: (i) DT-0: no demographic transparency; (ii) DT-1: demographic transparency in the form of charts; and (iii) DT-2: demographic transparency in the form of charts as well as individual photos accompanying each data subject. Therefore, a manipulation check is required to test whether the manipulation was successful. “manipulation checks are designed to ensure that subjects have, indeed, been manipulated as intended, a validity that can be empirically determined” (Boudreau et al. 2001, p. 5). To that end, participants were asked to respond to three questions about the level of demographic transparency of the subjects of their decision (Appendix A) and the responses were analyzed using analysis of variance (ANOVA). The results of the ANOVA test, depicted in Table 5.5, show that the treatment for demographic transparency was successful.

In addition, looking at the Post Hoc Mean Comparisons, we can compare whether the level of perceived DT was significantly increased from DT-0 to DT-1 and from DT-1 to DT-2. The results confirmed expectations as the level of DT is higher for participants in group DT-2 than participants in group DT-1 (M=5.64 versus 4.64, $\rho<.001$). Similarly, participants in group DT-1 perceived higher demographic transparency than participants in DT-0 (M=4.64 versus 3.82, $\rho<.001$).

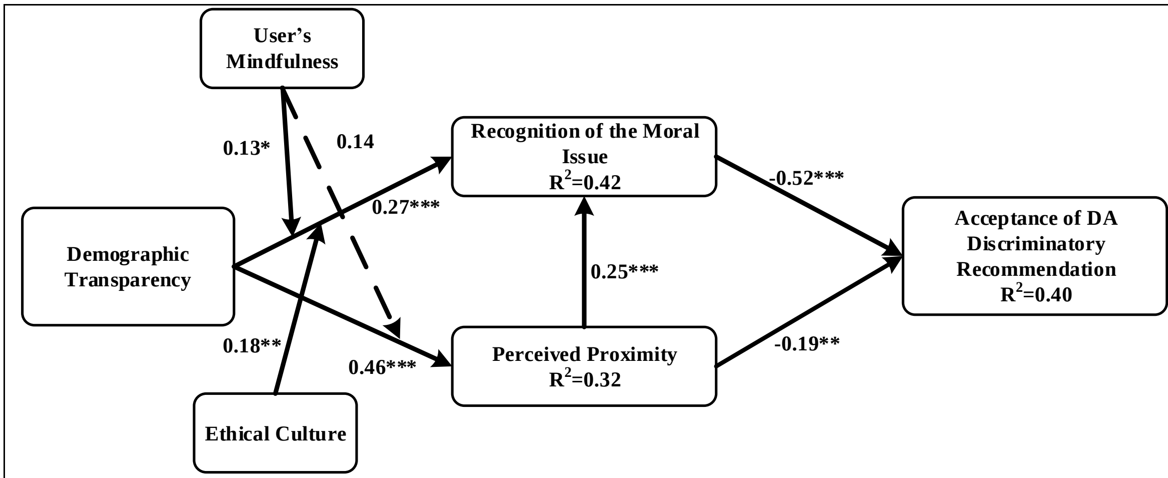
Table 5.5. One-Way ANOVA Analysis for DT Manipulation Check

Level of DT	Mean	Std. Deviation	Std. Error	95% Confidence Interval		ANOVA (Between Groups)			
				Lower Bound	Upper Bound	Sum of Squares	Mean Square	F	Sig.
DT-0	3.82	1.57	0.20	3.42	4.23	100.70	50.35	29.90	0.000
DT-1	4.64	1.37	0.17	4.30	5.00				
DT-2	5.64	0.84	0.11	5.42	5.85				

5.2.3. Results of Hypotheses Testing

Having established the appropriateness of the reliability and validity of the constructs as well as having ruled out the presence of common method bias, the next step is to provide evidence for the proposed theoretical hypotheses by examining the structural model. To that end, Structural Equation Modeling using SmartPLS 3.0 was employed to evaluate the proposed hypotheses and the significance of the path coefficients.

The results, shown in Figure 5.1 and Table 5.6, indicate that DT positively influenced recognition of the moral issue (β=0.27; ρ<0.001) and perceived proximity (β=0.46; ρ<0.001), supporting H1 and H2. Mindfulness positively moderates the impact of DT on recognition of the moral issue (β=0.13; ρ<0.05), supporting H3a, but did not influence the relationship between DT and proximity (β=0.14; ρ>0.05), thus H3b was not supported. Organizational ethical culture positively moderates the relationship between DT and recognition of the moral issue (β=0.18; ρ<0.005), which supports H4. Recognition of the moral issue negatively influenced acceptance of the recommendation (β=-0.52; ρ<0.001), which supports H5. Perceived proximity positively influenced recognition of the moral issue (β=0.25; ρ<0.001), supporting H6, and negatively influenced acceptance of the recommendation (β=0.19; ρ<0.005), thus supporting H7.



***: Significant (ρ <0.001)
**: Significant (ρ <0.005)
*: Significant (ρ <0.05)
Dotted line: insignificant (ρ>0.05)

Control Variables:
Gender, Age, Moral Identity, Impression Management, Prior knowledge of the possibility of DA tools generating discriminatory recommendations

Figure 5.1. PLS Model Results

Table 5.6. Validation of the Study Hypotheses				
Hypothesis	**Path coefficient**	**t-statistic**	**Significance**	**Supported?**
H1: In the context of a potentially discriminatory data analytics recommendation, DA tools providing higher demographic transparency will increase the likelihood of users' recognition of the moral issue.	0.268	3.782	0.000	Yes
H2: In the context of a potentially discriminatory data analytics recommendation, DA tools providing higher demographic transparency will increase users' perceptions of proximity toward the subjects of their decision.	0.458	8.835	0.000	Yes
H3a: In the context of a potentially discriminatory data analytics recommendation, users' mindfulness moderates the relationships between demographic transparency and recognition of the moral issue, such that the effect is stronger for individuals higher in mindfulness than for those lower in mindfulness.	0.129	2.328	0.020	Yes
H3b: In the context of a potentially discriminatory data analytics recommendation, users' mindfulness moderates the relationships between demographic transparency and perceived proximity towards the subjects of their decision, such that the effect is stronger for individuals higher in mindfulness than for those lower in mindfulness.	0.137	0.776	0.438	No

Table 5.6. Validation of the Study Hypotheses

Hypothesis	Path coefficient	t-statistic	Significance	Supported?
H4: In the context of a potentially discriminatory data analytics recommendation, Ethical culture of the organization moderates the relationship between demographic transparency and recognition of the moral issue, such that the effect is stronger for individuals from organizations with stronger ethical cultures.	0.176	3.267	0.001	Yes
H5: In the context of a potentially discriminatory data analytics recommendation, users' recognition of the moral issue is negatively associated with their acceptance of the system's recommendation.	-0.522	7.891	0.000	Yes
H6: In the context of a potentially discriminatory data analytics recommendation, users who perceive more proximity toward the subjects of their decision are more likely to recognize the moral aspect of the issue.	0.25	3.73	0.000	Yes
H7: In the context of a potentially discriminatory data analytics recommendation, users who perceive more proximity toward the subjects of their decision are less likely to accept a DA discriminatory recommendation.	-0.188	2.912	0.004	Yes

5.2.4. Analyses of R-Squared and Effect Sizes

In addition to examining the strength and significance of the hypothesized relationship, the coefficient of determination values (aka R^2) of the endogenous constructs were calculated using SmartPLS 3.0. R^2 measures the proportion of the variance of the dependent variable that is explained by the independent variables (Gefen et al. 2000). Although no cut-off value has been established for this measure, large R^2 values are generally sought after. Falk and Miller (1992) state that R^2 of all endogenous constructs should be at least 0.10. However, Chin (1998) suggest that R^2 values of approximately 0.670 are considered substantial, values around 0.333 are moderate, and values around 0.190 are weak (Urbach and Ahlemann 2010). As illustrated in Figure 5.1, the R^2 obtained for almost all endogenous constructs in this study are above the suggested moderate value of 0.333. More specifically, the values of R-squared for constructs recognition of the moral issue, perceived proximity, and acceptance of potentially discriminatory recommendation are 0.42, 0.32, and 0.40 respectively.

Effect size (f^2) is used to evaluate the impact that an antecedent (independent) construct has on a dependent construct (Cohen 1988). More specifically, to determine if the predictor (independent) construct has a small, medium or large effect size on the criterion (dependent) construct, the values of 0.02 (small), 0.15 (medium), and 0.35 (large) were used (Roldán and Sánchez-Franco 2012). SmartPLS 3.0 was used to determine the effect size of the relationships hypothesized in this study and the results are depicted in Table 5.7. As can be seen in the table the effect sizes are varied (6 small and 2 large).

Table 5.7. Effect Sizes Analysis		
Relation	f^2	**Effect Size**
DT → Recognition of the moral issue	0.094	Small
DT → Perceived Proximity	0.307	Large
DT × Mindfulness → Recognition of the moral issue	0.029	Small
DT × Mindfulness → Perceived proximity	0.029	Small
DT × Ethical Culture → Recognition of the moral issue	0.054	Small
Perceived proximity → Recognition of the moral issue	0.074	Small
Recognition of the moral issue → Acceptance of the recommendation	0.359	Large
Perceived proximity → Acceptance of the recommendation	0.047	Small

5.2.5. Goodness of Fit of the Research Model

To evaluate the structural model proposed in this study, the Goodness of Fit (GoF) index was employed. The GoF index is defined as the "geometric mean of the average communality and average R^2 for all endogenous constructs" (Akter et al. 2011, p. 3) and is calculated as follows:

$$GoF = \sqrt{\frac{\sum_n AVE_n}{n} \times \frac{\sum_m R_m^2}{m}}$$

Where n is the number of total constructs and m is the number of endogenous constructs. Drawing on the above formula, the GoF value obtained for the proposed model in this study was 0.54, which far exceeds suggested threshold of 0.36 and thus indicates a good performance of the model (Wetzels et al. 2009).

5.2.6. Effect of the Different Levels of DT on Recognition of the Moral Issue and Perceived Proximity

A MANOVA test was conducted to examine the effects of the three levels of demographic transparency on recognition of the moral issue and perceive proximity. MANOVA test statistics included Pillai's trace, Wilks' lambda, Hotelling's trace, and Roy's largest root. The p-values of these statics were found to be significant ($\rho<0.001$). Therefore, to separately test the impacts of demographic transparency on recognition of the moral issue and perceived proximity, two one-way ANOVAs were conducted, the results of which are presented in Table 5.8. As can be seen in the table, the results of the ANOVA tests confirm the results of the SEM analysis as the impact of increasing demographic transparency on both recognition of the moral issue and perceived proximity are significant.

Table 5.8. ANOVA Summary Table for Recognition of the Moral Issue and Perceived Proximity					
Dependent Variable	**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Recognition of the moral issue	95.43	2	47.72	20.47	0.000
Perceived proximity	46.34	2	23.17	26.68	0.000

In order to test the difference in recognition of the moral issue among various levels of demographic transparency, group comparisons were conducted. The results are reported in Table 5.9. The results indicate that both DT-1 and DT-2 were observed to have significantly higher levels of recognition of the moral issue than DT-0. However, the increase in the mean of recognition of the moral issue from DT-1 to DT-2 is not statistically significant.

Similarly, group comparisons were employed to test whether significantly different levels of proximity exist among the groups experiencing various levels of demographic transparency. The results, as depicted in Table 5.10, indicate that both DT-1 and DT-2 have significantly higher levels of perceived proximity compared to DT-0. In addition, the mean of perceived proximity is significantly higher in DT-2 compared to DT-1.

Table 5.9. Results on Recognition of the Moral Issue: Multiple comparisons of Demographic Transparency						
(I) group	**(J) group**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Confidence Interval**	
					Lower bound	**Upper bound**
DT-0 (Mean: 3.9)	DT-1	-1.46(*)	0.28	0.000	-2.13	-0.79
	DT-2	-1.60(*)	0.28	0.000	-2.26	-0.93
DT-1 (Mean: 5.3)	DT-0	1.46(*)	0.28	0.000	0.79	2.13
	DT-2	-0.14	0.28	1.000	-0.80	0.53
DT-2 (Mean: 5.5)	DT-0	1.60(*)	0.28	0.000	0.93	2.26
	DT-1	0.14	0.28	1.000	-0.53	0.80

Table 5.10. Results on Perceived Proximity: Multiple comparisons of Demographic Transparency						
(I) group	**(J) group**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Confidence Interval**	
					Lower bound	**Upper bound**
DT-0 (Mean: 4.4)	DT-1	-0.49(*)	0.17	0.013	-0.89	0.08
	DT-2	-1.22(*)	0.17	0.000	-1.63	-0.82
DT-1 (Mean: 4.9)	DT-0	0.49(*)	0.17	0.013	0.08	0.89
	DT-2	-0.74(*)	0.17	0.000	-1.14	-0.33
DT-2 (Mean: 5.6)	DT-0	1.22(*)	0.17	0.000	0.82	1.63
	DT-1	0.74(*)	0.17	0.000	0.33	1.14

5.2.7. Interaction Plots for the Impact of User's Mindfulness and Ethical Culture

The significance of the moderating impacts of user's mindfulness and their organization's ethical culture on the relationship between demographic transparency and recognition of the moral issue as well as perceived proximity toward the subjects of the decision was tested through SEM methodology. The results show that user's mindfulness and their organization's ethical culture influence the impact of increasing the level of demographic transparency on recognition of the moral issue. However, the influence of user's mindfulness on increasing the level of perceived proximity as a result of exposure to higher levels of demographic transparency was non-significant. In order to further examine the above moderating effects, the Interaction software package was used to draw the interaction plots and help us better understand these moderating effects. The resulting plots are shown in Figures 5.2, 5.3, and 5.4. The values on each line, represent the unstandardized regression coefficient of the dependent variable on the independent variable at the level of the moderator for that specific interaction line.

As can be seen in Figure 5.2, on average, for users with mindfulness levels above (μ-σ), a positive significant relationship exists between increasing the level of demographic transparency and recognition of the moral issue at hand. However, this relationship, although positive, is not significant for users with mindfulness level (μ-2σ) and below.

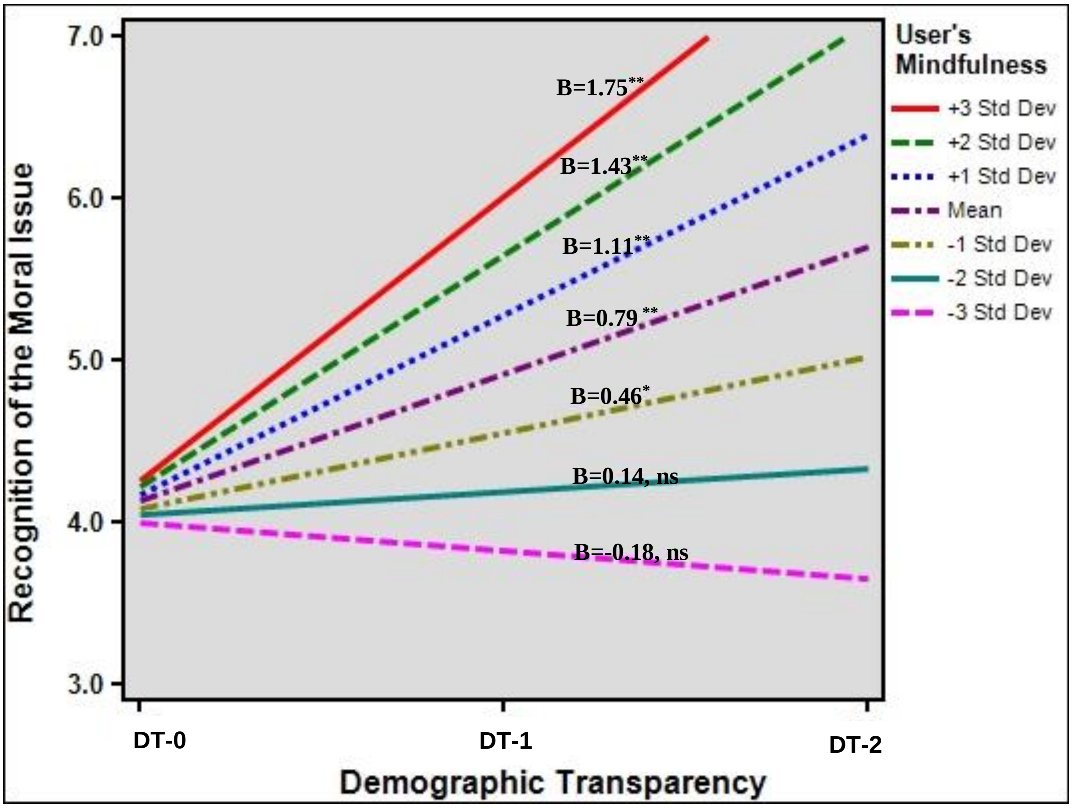


Figure 5.2. Interaction Plot for the Impact of User's Mindfulness on the Relation DT-Recognition of the Moral Issue

Figure 5.3 depicts the moderating influence of user's organizational ethical culture on the relationship between demographic transparency and recognition of the moral issue. As can be seen on the figure, for users with organizational ethical culture above (μ-σ), a positive significant relationship exists between increasing the level of demographic transparency and recognition of the moral issue at hand. However, this relationship is not significant for users with organizational ethical culture levels (μ-2σ) and below.

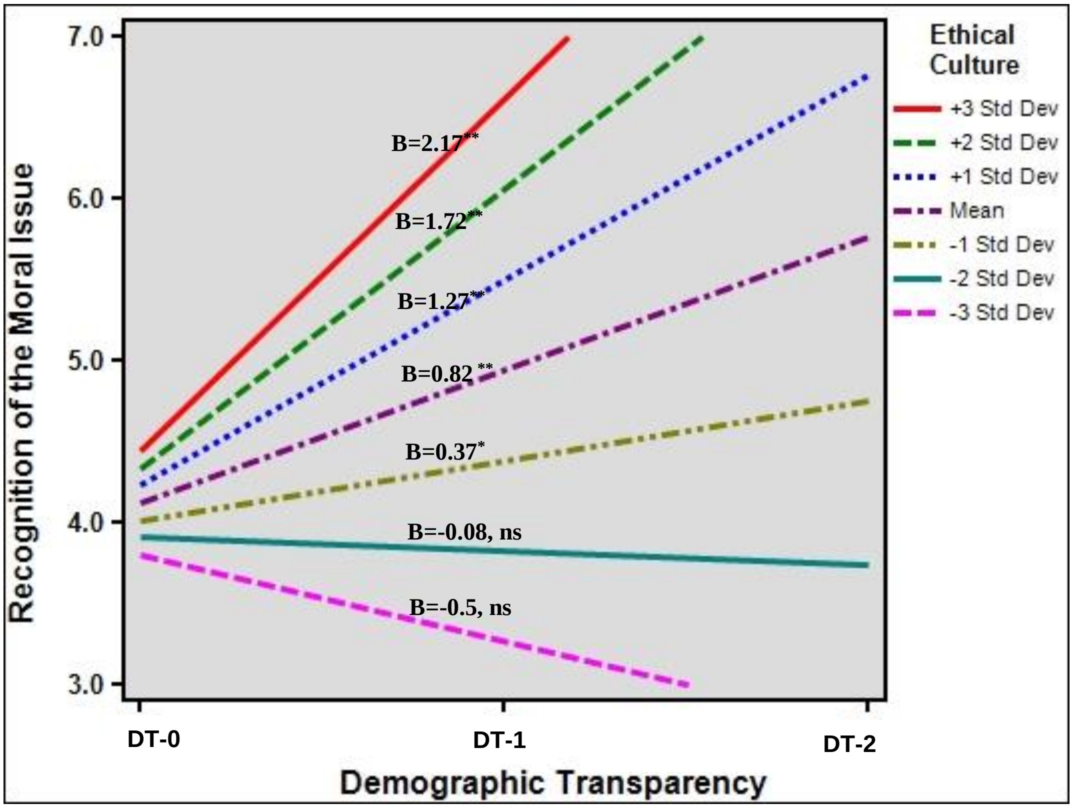


Figure 5.3. Interaction Plot for the Impact of Ethical Culture on the Relation DT-Recognition of the Moral Issue

Finally, Figure 5.4 depicts the moderating influence of user's mindfulness on the relationship between demographic transparency and perceived proximity toward the subjects of the decision. As can be seen in the figure, this relationship is significant for all levels of user's mindfulness. Therefore, user's mindfulness does not interact with demographic transparency to predict the level of perceived proximity. However, since perceived proximity is consistently higher for users with higher levels of mindfulness, a direct relationship between these two factors could be expected. This relationship will be tested in the section on saturated model analysis below.

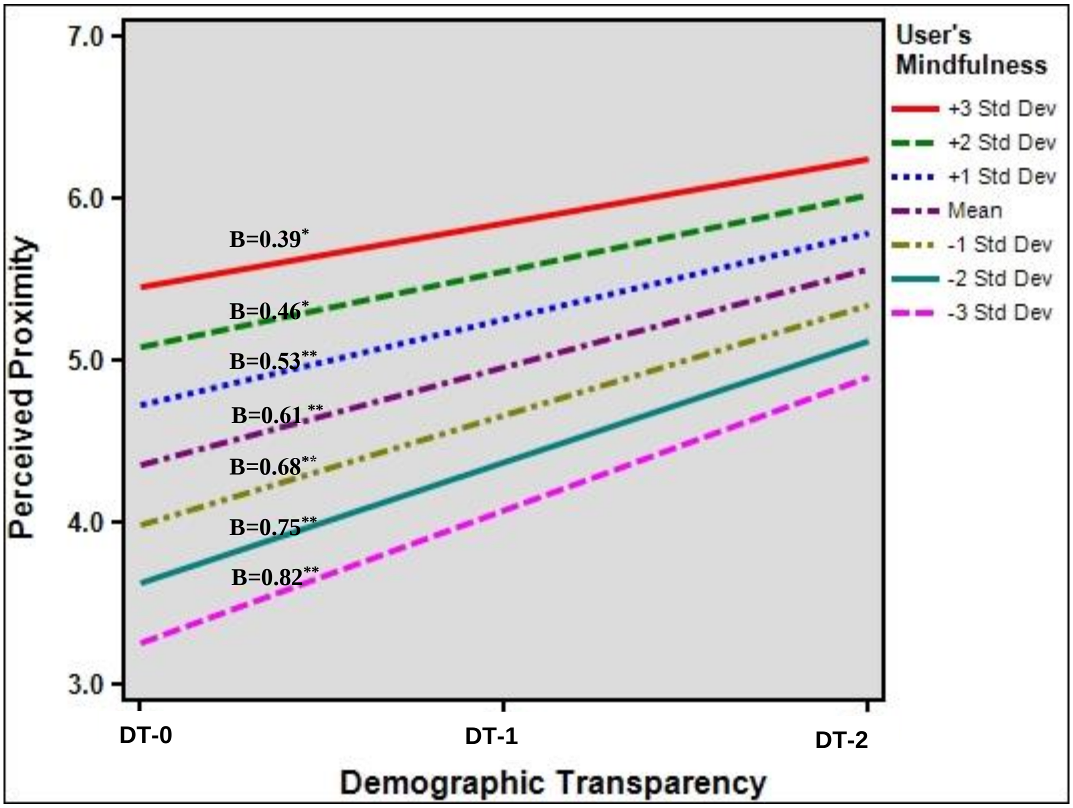


Figure 5.4. Interaction Plot for the Impact of User's Mindfulness on the Relation DT-Perceived Proximity

5.2.8. Analysis of the Impacts of Control Variables

As discussed earlier in Chapter 3, in addition to the measurement items included in the model, measurement items about a series of control variables were included in the questionnaire. Those variables were analyzed to control for their potential impact on the endogenous constructs in the research model. In total, 6 control variables were analyzed: users' gender, age, internalization and symbolization subscales of moral identity, prior knowledge of the possibility of data analytics tools generating discriminatory recommendations, and impression management. To analyze the impact of these variables, each was added to the model one at a time, by linking it to each endogenous variable in the model. Employing SmartPLS 3.0, the strength and significance of those paths were determined. The results are reported in Table 5.11 showing that none of these control variables except for the internalization dimension of moral identity impacted any of the endogenous constructs of the model significantly. Despite significant impacts of the internalization dimension of moral identity on recognition of the moral issue and acceptance of potentially discriminatory DA recommendation, when added to the model, none of the hypothesized paths in the model changed algebraic sign, nor did any of the paths become non-significant. Therefore, it can be concluded that the control variables did not alter the conclusions derived from the hypotheses of this study.

Table 5.11. Results on Control Variable Analysis			
Control Variable	**Endogenous Construct**	**Path Coefficient**	**Significance**
Gender (1=Female, 2=Male)[9]	Recognition of the Moral Issue	-0.06	n.s.
	Perceived Proximity	-0.03	n.s.
	Acceptance of the Recommendation	0.07	n.s.
Age	Recognition of the Moral Issue	-0.09	n.s.
	Perceived Proximity	-0.02	n.s.
	Acceptance of the Recommendation	0.02	n.s.
Moral Identity (Internalization)	Recognition of the Moral Issue	0.16	0.005
	Perceived Proximity	0.1	n.s.
	Acceptance of the Recommendation	-0.11	0.04
Moral Identity (Symbolization)	Recognition of the Moral Issue	0.09	n.s.
	Perceived Proximity	0.08	n.s.
	Acceptance of the Recommendation	0.05	n.s.
Prior knowledge of the possibility of DA tools generating discriminatory recommendations	Recognition of the Moral Issue	0.11	n.s.
	Perceived Proximity	0.03	n.s.
	Acceptance of the Recommendation	0.01	n.s.
Impression Management	Recognition of the Moral Issue	0.05	n.s.
	Perceived Proximity	0.02	n.s.
	Acceptance of the Recommendation	0.01	n.s.

[9] Although the questionnaire included items for Genders other than female and male (e.g., transgender female and transgender male), the received responses did not include those options.

5.2.9. Saturated Model Analysis

To explore possible non-hypothesized relationships among the variables of the research model, a saturated model was created by establishing all possible links among the variables in the originally proposed research model. Subsequently, SmartPLS 3.0 was employed to perform path analysis. Results of this analysis are depicted in Table 5.12.

As can be seen in Table 5.12, the direct relationship between demographic transparency and approval of the potentially discriminatory recommendation is insignificant. Therefore, it was deemed relevant to examine whether this relationship is fully mediated by recognition of the moral issue and perceived proximity. To that end, I utilized the four-step procedure proposed by Baron and Kenny (1986), who suggest that to do a mediation analysis the following four steps should be taken: (1) regress the potential mediator on the independent variable; (2) regress the dependent variable on the independent variable; (3) regress the dependent variable on both the independent variable and the potential mediator; and (4) examine the coefficient of the independent variable from the previous steps. Therefore, as the first step, the impacts of DT on recognition of the moral issue and perceived proximity were calculated (DT → Recognition of the moral issue: β=0.39, ρ<0.001, and DT → Perceived proximity: β=0.47, $\rho < 0.001$). Then the direct effect of demographic transparency on approval of the potentially discriminatory recommendation was tested, which was significant (β= -0.36, $\rho < 0.001$). Next, when recognition of the moral issue and perceived proximity were added to the model together with demographic transparency to predict approval of the recommendation, the effect of demographic transparency on approval of the potentially discriminatory recommendation was no longer significant with a coefficient of -0.09 ($\rho > 0.05$). Thus, recognition of the moral issue and perceived proximity fully mediate the relationship between demographic transparency and approval of the DA discriminatory recommendation.

Table 5.12. PLS Results on Non-Hypothesized Paths- Saturated Model Analysis

Non-Hypothesized Path	Path Coefficient	Significance	Validation
DT → Approval of the recommendation	-0.09	0.18	Rejected
User's mindfulness → Recognition of the moral issue	0.09	0.13	Rejected
User's mindfulness → Perceived proximity	0.25	0.000	Supported
User's mindfulness → Approval of the recommendation	-0.05	0.44	Rejected
Ethical culture → Recognition of the moral issue	0.3	0.000	Supported
Ethical culture → Perceived proximity	0.04	0.51	Rejected
Ethical culture → Approval of the recommendation	0.05	0.43	Rejected

Furthermore, the direct impact of the two moderators (i.e., user's mindfulness and ethical culture) on the three endogenous constructs were examined. Whereas the impact of user's mindfulness on recognition of the moral issue and acceptance of the recommendation is insignificant, user's mindfulness significantly increases the user's perceived proximity toward the subjects of his/her decision. In addition, the direct impacts of the level of ethical culture in user's organization on recognition of the moral issue is significant. However, the level of ethical culture does not significantly impact perceived proximity and acceptance of the recommendation.

5.2.10. Supplementary Analysis on Acceptance of the Recommendation across the Three Treatment Groups

As discussed earlier, the results of the SmartPLS analysis shows that in the absence of the two mediating variables (i.e., recognition of the moral issue and perceived proximity), the direct impact of demographic transparency on acceptance of the recommendation is significant (β= -0.36, $\rho < 0.001$). In order to further examine the differences among the three levels of demographic transparency in terms of impacting the acceptance of the discriminatory recommendation, chi-square test of independence was used (Daniel 1990). In line with the result obtained from SmartPLS, the results of the chi-square analysis confirmed that demographic transparency has a significant relationship with acceptance of the potentially discriminatory recommendation ($\chi^2(2)$=24.2, $\rho < 0.001$). The breakdown of the number of participants in each treatment group (i.e., DT-0, DT-1, and DT-2) and their acceptance/non-acceptance of the DA recommendation is provided in Table 5.13.

Table 5.13 Frequency counts for chi-square analysis of the relation DT-acceptance of potentially discriminatory DA recommendation				
Independent Variable		**DA Recommendation Not Accepted**	**DA Recommendation Accepted**	**Total**
Demographic Transparency	DT-0	16	45	61
	DT-1	32	29	61
	DT-2	43	18	61
Total		91	92	183

To shed more light on whether there is a significant different between different levels of demographic transparency and acceptance of the potentially discriminatory recommendation, two separate chi-square tests of independence were performed. The results confirmed that acceptance of the recommendation was significantly lower among participants in treatment DT-1 than participants in DT-0 ($\chi^2(1)=8.8$, $\rho < 0.005$) and among participants in treatment DT-2 than participants in DT-1 ($\chi^2(1)=4.1$, $\rho < 0.05$).

5.2.11. Supplementary Analysis on Acceptance of Potentially Discriminatory Recommendation Using Logistic Regression

As discussed previously in this thesis, SmartPLS has been used to test the relationships *recognition of the moral issue - acceptance of potentially discriminatory DA recommendation* and *perceived proximity - acceptance of potentially discriminatory DA recommendation*. However, due to some concerns in regard to using SmartPLS for testing relationships with an ultimate dependent variable measured using a categorical scale (Hair et al. 2017) which is the case in this study, the above relationships were also tested through logistic regression method and by employing IBM SPSS Statistics 24, the results of which are presented in this section.

Table 5.14 depicts the results of the logistic regression analysis in terms of various metrics for testing the model fit. According to the Hosmer-Lemeshow goodness-of-fit test, the model was significantly better at determining the probability of acceptance of the recommendation than random chance. In addition the high value of the two pseudo R-squares (i.e., 0.375 and 0.5) also testify to a good fit of the model to the data.

Table 5.14 Analysis of acceptance of potentially discriminatory recommendation results for model fit		
Cox & Snell R Square	**Nagelkerke R Square**	**Hosmer and Lemeshow Test (chi-square, p <)**
0.375	0.5	(15.1, 0.06)

Table 5.15 shows the results of the logistic regression analysis in terms of the extent and significance of each independent variable on the outcome variable. As can be seen both variables (i.e., recognition of the moral issue and perceived proximity) have a significant impact on acceptance of potentially discriminatory recommendation.

Table 5.15 Logistic Regression Results							
Independent Variable	**B**	**SE**	**Wald**	**df**	**Sig.**	**Odds Ratio**	**95% CI for Odds Ratio**
Recognition of the moral issue	-0.898	0.154	34.2	1	<0.001	0.41	(0.3,0.55)
Perceived proximity	-0.54	0.218	6.12	1	<0.05	0.583	(0.38,0.89)

It is noteworthy though that regression coefficients in the above table are unstandardized. Therefore, the following formula proposed by Menard (1995) was used to compute the standardized coefficients. The standardized regression coefficients for the impacts of recognition of the moral issue and perceived proximity on acceptance of potentially discriminatory DA recommendation were computed as -0.53 and -0.20 respectively. These values are also in line with the outcome of the SmartPLS analysis which had provided us with -0.52 and -0.19 for the two relationships. In addition, the formula suggested by Menard (1995) was employed to compute the exact R^2 of the endogenous construct, acceptance of potentially discriminatory DA recommendation and yielded an a result of 43%. This value is very close to the output of SmartPLS discussed previously (i.e., 40%). In summary, it can be concluded that the results of the SmartPLS analysis for the impacts of the two mediating variables on the ultimate endogenous variable in the research model were confirmed by the logistic regression analysis.

DISCUSSIONS

Data Analytics tools are increasingly being used to make organizational decisions. However, there have been societal concerns raised about such use of data analytics tools as, in some circumstances, these tools have been shown to provide decision makers with discriminatory recommendations (Barocas and Selbst 2016; Gangadharan et al. 2014; Newell and Marabelli 2015). To date several researchers have investigated the technical aspects of this problem and suggested methods that can help decrease the likelihood of generating such recommendations. However, to the best of my knowledge, no study has investigated the human aspect of decision-making using data analytics tools and whether and how decision makers can be equipped with tools that help them recognize the discriminatory nature of the recommendation (if any) provided by a DA tool.

To that end, this study suggests providing data analytics users with demographic transparency as a specific type of decisional guidance and investigates the impact of such provision on users' recognition of the moral issue, their perceived proximity toward the subjects of the decision, and ultimately, not accepting the DA discriminatory recommendation. The experimental results confirm that demographic transparency is indeed an effective tool in the context of a potential discriminatory recommendation.

The confirmed effect of increasing demographic transparency on recognition of the moral issue ($\beta = 0.27$, $\rho < 0.001$) is in line with the cognitive elaboration model of ethical decision-making, suggesting that increasing the ability and motivation of individuals can affect their recognition of the moral aspect of the decision they are asked to make.

Demographic transparency provides aggregated demographic information about the proportion of each demographic class (e.g., female and male) in its pertinent demographic category (e.g., sex) in the original full dataset and the DA-recommended sample and thereby enables decision makers to recognize traces of potential discrimination (if any) in the recommendation of the DA tool. Furthermore, demographic transparency by providing more detailed information about and emphasizing the human aspects of the subjects of the decision, increases the level of proximity decision makers feel toward the subjects of their decision ($\beta = 0.46$, $\rho < 0.001$).

In line with several previous studies that have observed the impact of recognizing the moral aspect of an issue on making an ethical decision (Fleischman and Valentine 2003; Haines et al. 2008), this study also found a significant negative relationship between recognition of the moral issue and approval of the discriminatory recommendation ($\beta = -0.52$, $\rho < 0.001$). In addition, the hypothesis regarding the negative impact of perceived proximity on accepting the discriminatory recommendation was supported ($\beta = -0.19$, $\rho < 0.005$). Similar results in terms of the negative impact of perceived proximity on committing an unethical behavior have been observed in previous studies (e.g., Carlson et al. 2002; Karacaer et al. 2009; Leitsch 2006).

Delving deeper into the impact of demographic transparency on recognition of the moral issue in the context of a potentially discriminatory data analytics recommendation showed that providing aggregated demographic information about the subjects of the decision in the form of charts increases the level of recognition of the moral issue at hand. This result is in line with some common practices in identifying instances of potential discrimination (e.g., the extended ratio measure as discussed in Chapter 2). These practices aim at producing statistical parity and group fairness (Chouldechova 2017). In addition, in line with construal level theory and the literature on moral intensity, demographic transparency information in the form of charts that depict the proportion of each demographic class in its demographic category successfully enhanced user's perceived proximity toward the subjects of the analysis and decision.

Increasing the level of demographic transparency by providing individual-level demographic information in the form of photos mainly decreased the acceptance of the potentially discriminatory recommendation through increasing the perceived proximity. The impact of such individual-level information on increasing recognition of the moral issue turned out to be insignificant. I believe that the reason for the insignificant relation is that in the experiment, the participants who received the individual level information about the data subjects (i.e., photos) were also provided with aggregated demographic information in the form of charts, which also depicted the disproportionate lower ratio of selected females compared to males. Had I only provided the individual-level information, the results might have been different.

In addition, the results of the experiment confirm that user's mindfulness positively impact the effects of demographic transparency on recognition of the moral issue making this relationship stronger at higher levels of mindfulness ($\beta = 0.13$, $\rho < 0.05$). This is in line with the existing literature on the notion of mindfulness, which asserts that individuals with higher levels of mindfulness tend to be more aware of and pay more attention to what is taking place at the moment (Brown and Ryan 2003) and therefore, are more likely to recognize the moral aspects of issues (Ruedy and Schweitzer 2010). Similarly, participants with higher levels of mindfulness were observed to be more likely to attend to the demographic transparency information provided to them and as a result recognize the potential unethical discriminatory dimension of the generated recommendation.

However, the hypothesis about the moderation impact of user's mindfulness on the impact of DT on perceived proximity was not supported ($\beta = 0.14$, ns). One reason might be the salience of the photos of the individuals, which were the main avenue for increasing the perceived proximity. Since the number and size of these photos were noticeable, participants with all levels of mindfulness noticed and processed them. It is noteworthy that, although not originally hypothesized, a positive direct relationship was found between a user's mindfulness and their perceived proximity toward the subjects of their decision ($\beta = 0.25$, $\rho < 0.001$). I believe the reason lies in what Barnes et al. (2007) call Flexibility of awareness and attention. They assert that "Like a zoom lens, [mindfulness] can move back from particular states of mind to gain a larger perspective on what is taking place (clear awareness), and can also zero in on situational details (focused attention) according to inclination or circumstance" (Barnes et al. 2007, p. 213).

In addition to user's mindfulness, the ethical culture of the organization an individual is employed at was shown to enhance the positive relationship between the provision of demographic transparency information to the individual and their recognition of the moral issue related to a potentially discriminatory recommendation provided by a DA tool ($\beta = 0.18$, $\rho < 0.001$). This is in agreement with the existing business ethics literature that suggest that an ethical culture in an organization influences individuals' beliefs and behavior and delineates what behaviors are/are not considered acceptable (Douglas et al. 2001; Sweeney et al. 2010; Treviño et al. 1998; Zhang et al. 2009).

In sum, this study empirically shows that providing aggregated and individual demographic information decreases the likelihood of acceptance of a potentially discriminatory DA recommendation and thereby makes important contributions to theory and provides essential implications for the practice which are discussed next.

5.3. Contributions to the Theory and Research

The main goal in this research is to devise a method to alleviate the problem of discriminatory decision making when using DA tools. To that end, this study conceptualized and operationalized the notion of demographic transparency as a means of providing DA users with demographic information about the subjects of their DA-aided decisions. This study stands to make significant contributions to theory. First, to the best of my knowledge, this is the first empirical study that examines the issue of discriminatory decision making using DA tools and strives to ameliorate the problem by focusing on the human aspects of decision making in using such tools. Previous studies have either produced generalized suggestions that using those tools can lead to making discriminatory decisions or merely focused on the technical antecedents of such discriminatory recommendations. Therefore, this study advances the data analytics literature by proposing the notion of demographic transparency as a form of decisional guidance to address the issue by focusing on the cognition and attitude of the human decision maker.

This study addressed the impact of demographic transparent DAs relative to non-demographic transparent DAs in terms of users' recognition of the moral issue in the context of a DA potentially discriminatory recommendation as well as users' perceived proximity toward the subjects of their decision. It was demonstrated that provision of aggregated demographic information about different demographic classes and their selection rates as recommended by the DA tools with the demographic transparent DA, not only increases user's recognition of the potential moral issue but also enhances their perceived proximity toward the subjects of their decision. These in turn lead to a lower acceptance of a potentially discriminatory recommendation. Therefore, demographic transparent DAs are capable of promoting group fairness, which has been suggested as an important concern in the literature on the ethics of big data analytics (Barocas and Selbst 2016).

In addition, grounded in the cognitive elaboration model of ethical decision-making and the moral intensity literature, this study demonstrated that providing individual level demographic information in the form of photos can decrease the readily acceptance of a potentially discriminatory recommendation generated by a DA tool. This is important in that some recent conceptual and technical papers (e.g., Williams et al. 2018; Žliobaitė and Custers 2016) have also recognized the importance of collecting individual demographic information and using it in the analysis by DA tools to decrease the likelihood of making a discriminatory decision. To the best of my knowledge, this is the first empirical study to examine the role of provision of individual level demographic information in the context of discriminatory recommendation of data analytics tools. Thus, this study will help researchers better understand why providing individual demographic information about the data subjects in a DA context can reduce the likelihood of making a discriminatory decision.

In summary, the present study integrates three streams of research, data analytics literature, business ethics literature, and DSS guidance studies to be the first empirical study that reveals the impact of availability of aggregated and individual-level demographic information on DA users' perceptions and behavior. Providing such information as decisional guidance to enlighten DA users as they make their decisions based on the recommendations of these tools is also a new method to deal with the issue of information asymmetry between data analytics tools and their users (Mayer-Schönberger and Cukier 2013; Newell and Marabelli 2015).

This study also contributes to the social psychology literature by extending the issue-contingent model of ethical decision making (Jones 1991) to a new context of organizational decision making using Data Analytics tools. This model has been used to explain unethical decision-making and behavior in organizations (Leitsch 2006; Mencl and May 2009; Singhapakdi et al. 1996b). However, this study shows that this model is also useful in explaining technology-supported decision making and data analytics tools. More specifically, this study recommends a method to enhance the DA users' proximity to the subjects of their decisions, a dimension of the issue-contingent model of ethical decision-making. Previously proximity was primarily used and operationalized in terms of the victims being a co-worker, colleague, friend, family member, or community member of the decision maker (Frey 2000; Leitsch 2006; May and Pauli 2002; Morris and McDonald 1995; Paolillo and Vitell 2002; Singer et al. 1998; Singhapakdi et al. 1999; Watley and May 2004). However, drawing on the construal level theory, this study shows that the notion of proximity can also be manipulated and enhanced through providing aggregated demographic information about the population of decision subjects as well as individual-level information. Both these methods can be used in the context of technology-supported decision making and behavior and therefore, can be useful not only in the context of DA use but in regard to other uses of IT which has been long blamed for contributing to decreasing the interpersonal warmth (Cummings 2006; Haslam 2006).

5.4. Implications for Practice

While the preceding comments focus on theoretical developments, the results regarding the impact of demographic transparency on DA users' perceptions and behavior have significant practical implications for organizations as well. This study suggests a combination of situational, individual, and organizational factors that together can help reduce the incidences of unintentional discriminatory decisions when using data analytics tools. On the one hand, a demographically transparent DA enables users to more easily recognize whether a potential discrimination exists in a DA recommendation and take steps to investigate whether such discrimination indeed exists[10]. This is especially important as in the age of big data, decision makers in order to cope with the complexity associated with such large data sets completely rely on data analytics tools (Newell and Marabelli 2015). In such circumstances DA users have no alternative but to readily accept DA recommendations trusting them to be accurate and non-discriminatory when in fact that might not be the case. Thus, having access to the kind of decisional guidance recommended in this study provides DA users with a chance to flag potential discriminatory recommendations and, before making a final decision, pause to explore whether there is indeed a discriminatory aspect to the recommendation at hand.

On the other hand, the demographically transparent DA enhances users' perception of proximity toward the subjects of their decision. Together these two variables decrease the likelihood that users will readily accept a DA recommendation that is potentially discriminatory against a group of data subjects without investigation. Thus, practitioners, particularly those who are concerned about the issue of unintentional discriminatory decision making by their DA users, are advised to incorporate the demographic transparency functions into their DA tools.

[10] As discussed previously in this thesis, a situation that due to over- or under-representation of a demographic group, seems potentially discriminatory, might not necessarily be discriminatory as there might be some business necessities behind the over- or under-representation of individuals from that particular demographic group.

It is noteworthy that in some cases and due to legal or privacy concerns, organization are reluctant to provide their DA users with individual-level demographic information about data subjects. In such circumstances, aggregated-level demographic transparency decisional guidance exhibits can be employed. The results of this study show that such exhibits, which visually depicts the aggregated demographic information about the proportion of members of each demographic class in both the original pool and the recommended sample by the data analytics tool leads to significantly more users recognizing the moral issue as well as perceiving more proximity toward the data subjects leading them to be less accepting of potentially discriminatory recommendations. This level of demographic transparency although not as effective as providing both the aggregated and individual-level demographic information, can still significantly decrease the likelihood of accepting a potentially discriminatory recommendation.

This study also shows the important role of individual users' mindfulness in helping practitioners who seek to lower the level of discriminatory decisions made by their DA users. Mindfulness was shown to not only enhance the impact of demographic transparency information on recognition of the moral issue but also to be directly associated with the level of proximity one feels toward the subjects of their decisions. As a result, organization should employ and invest in users who are high on this trait. In addition, it is notable that mindfulness is not a stable trait but rather can be increased by training programs (Hayes et al. 1999; Kabat-Zinn 1982; Linehan 1993; Segal et al. 2002) which organizations should pursue.

The results also demonstrated the positive impact of organizational ethical culture on user's recognition of the ethical aspect of the decision they were asked to make. Since such recognition is the first step in triggering ethical thinking, decision making and behavior, organizations need to consider and invest in cultivating an ethical culture that encourages employees to engage in ethical decision making.

5.5. Limitations and Future Research

Notwithstanding its significance and contributions, several limitations exist for this study that provide avenues for future research. First, participants for this study were selected from North American middle managers. Given the potential impacts of culture on users' attitude toward IT use as well as moral behaviors, caution should be exercised in generalizing the results of this study to data analytics users in other geographic regions. To generalize these results, conducting additional studies with different subject demographics is necessary.

Second, this study was conducted in a context in which the participants use and evaluate demographic transparency for the first time. Therefore, further research is warranted to test the impact of demographic transparency on DA users when they are repeatedly exposed to it potentially developing tolerance to its effects.

Another limitation is that this study only focuses on discrimination against one demographic class (i.e., females). Future research is warranted to examine if the results of this study are generalizable to other demographic categories (e.g., age, race, and marital status).

Furthermore, though not a true limitation, the context of this study entails a situation about selecting a number of employees to be sent to a training program on leadership skills. Since participants were told that being sent to such a training program is a great experience which can help employees in receiving promotions in the future, the study includes selecting employees for the purpose of receiving a reward. Future research can examine whether the results of this study is generalizable to situations that entail a "punishment" (e.g., selecting employees to be laid off).

Another potential limitation relates to the fact that photos of individual subjects are not available in all contexts. Whereas such photos are available in situations related to internal organizational decisions (as was the case in this study), such photos might be unavailable or illegal to use in some other situations such as marketing-related analyses and decision-making. Nonetheless, using photos is only one means of increasing the perceived proximity toward other individuals. Therefore, future research should examine whether other approaches (e.g., avatars, short biographies, etc.) can have a similar impact in terms of increasing the level of perceived proximity.

Finally, in this study a quantitative research methodology was employed for which an online experiment was conducted. The results of this study demonstrated the effectiveness of the proposed approach and confirmed the majority of the proposed hypotheses. Future studies can employ qualitative methodologies such as undertaking interviews with DA users after being exposed to the demographic transparency decisional guidance. The results of such an approach can generate valuable insights in regard to subjects' cognitive processes towards evaluating the recommendation provided by the DA tool and the demographic transparency information.

5.6. Concluding Remarks

This study addressed an important gap in the DA literature in terms of providing DA users with informative decisional guidance to help them recognize traces of potential discrimination in the recommendations of a DA tool and thus, not readily accept them. To that end, this study proposed, conceptualized, and operationalized the notion of demographic transparency that aims at providing DA users with aggregated and individual demographic information about the data subjects of their decision. This study aimed at fulfilling two research objectives: (1) To employ a theoretical model to investigate how and to what extent does providing aggregated and individual demographic information regarding the human subjects of DA recommendations would reduce the incidence of users' acceptance of potentially discriminatory recommendations of DA systems. (2) To investigate how and to what extent do user's mindfulness and organizational ethical culture impact the relationship between providing aggregated and individual demographic information regarding the human subjects of DA recommendation and DA user's acceptance of potentially discriminatory recommendations of those systems. To that end, cognitive elaboration model of ethical decision-making, construal level theory, and the literature on issue-contingent moral intensity were employed as the theoretical underpinnings, to propose a research model comprising seven hypotheses. This study devised an online experimental approach using a fictitious data analytics tool developed for this study to subject the seven proposed hypotheses to quantitative empirical tests. Data were collected from 183 middle managers employed in organizations in North America. Analysis of data using several quantitative data analysis methodologies confirmed the majority of the expectations and demonstrated the effectiveness of the proposed demographic transparency methods to combat potentially discriminatory recommendations put forth by DA tools. To the best of my knowledge, this is the first empirical study on discriminatory decision making using data analytics tools. Thus, it can serve as a foundation for future empirical studies in the context of the ethics of data analytics usage in organizations.

REFERENCES

Akter, S., D'Ambra, J., and Ray, P. 2011. "An Evaluation of Pls Based Complex Models: The Roles of Power Analysis, Predictive Relevance and Gof Index," *17th Americas Conference on Information Systems (AMCIS2011)*, Detroit, USA, pp. 1-7.

Amit, E., Algom, D., and Trope, Y. 2009. "Distance-Dependent Processing of Pictures and Words," *Journal of Experimental Psychology: General* (138:3), pp. 400-415.

Amit, E., Algom, D., Trope, Y., and Liberman, N. 2008. "Thou Shalt Not Make Unto Thee Any Graven Image": The Distance-Dependence of Representation," in *The Handbook of Imagination and Mental Simulation*, K.D. Markman, W.M.P. Klein and J.A. Suhr (eds.). New York, NY: Psychology Press, pp. 53-68.

Aquino, K., and Reed, I. 2002. "The Self-Importance of Moral Identity," *Journal of personality and social psychology* (83:6), p. 1423.

Arakawa, L. L., and Sonen, M. P. 2009. "Caught in the Backdraft: The Implications of Ricci V. Destefano on Voluntary Compliance and Title Vii," *University of Hawaii Law Review* (32), pp. 463-483.

Australian Legislation. 2015. "(a) Age Discrimination Act, 2004; (B) Australian Human Rights Commission Act, 1986; (C) Disability Discrimination Act, 1992; (D) Racial Discrimination Act, 1975; (E) Sex Discrimination Act, 1984; (F) Victoria Equal Opportunity Act, 1995, (G) Queensland Anti-Discrimination Act, 1991." http://www.hreoc.gov.au.

Bagozzi, R. P., and Yi, Y. 1988. "On the Evaluation of Structural Equation Models," *Journal of the academy of marketing science* (16:1), pp. 74-94.

Bandura, A. 1999. "Moral Disengagement in the Perpetration of Inhumanities," *Personality and social psychology review* (3:3), pp. 193-209.

Bar-Anan, Y., Liberman, N., and Trope, Y. 2006. "The Association between Psychological Distance and Construal Level: Evidence from an Implicit Association Test," *Journal of Experimental Psychology: General* (135:4), pp. 609-622.

Barclay, D., Higgins, C., and Thompson, R. 1995. "The Partial Least Squares (Pls) Approach to Casual Modeling: Personal Computer Adoption Ans Use as an Illustration," *Technology Studies* (2:2), pp. 285-324.

Barnard, A. 2001. "On the Relationship between Technique and Dehumanization," in *Advancing Technology, Caring, and Nursing*, R. Locsin (ed.). Westport, CT: Greenwood, pp. 96-105.

Barnes, S., Brown, K. W., Krusemark, E., Campbell, W. K., and Rogge, R. D. 2007. "The Role of Mindfulness in Romantic Relationship Satisfaction and Responses to Relationship Stress," *Journal of marital and family therapy* (33:4), pp. 482-500.

Barnett, T. 2001. "Dimensions of Moral Intensity and Ethical Decision Making: An Empirical Study," *Journal of Applied Social Psychology* (31:5), pp. 1038-1057.

Barnett, V., and Lewis, T. 1974. *Outliers in Statistical Data*, (2nd ed.). New York: Wiley.

Barocas, S., and Selbst, A. D. 2016. "Big Data's Disparate Impact," *California Law Review* (104), pp. 671-732.

Baron, R. M., and Kenny, D. A. 1986. "The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology* (51:6), pp. 1173-1182.

Bishop, S. R. 2002. "What Do We Really Know About Mindfulness-Based Stress Reduction?," *Psychosomatic medicine* (64:1), pp. 71-83.

Boudreau, M.-C., Gefen, D., and Straub, D. W. 2001. "Validation in Information Systems Research: A State-of-the-Art Assessment," *MIS quarterly* (25:1), pp. 1-16.

Brown, K. W., and Ryan, R. M. 2003. "The Benefits of Being Present: Mindfulness and Its Role in Psychological Well-Being," *Journal of personality and social psychology* (84:4), pp. 822-848.

Burrell, J. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society* (3:1), pp. 1-12.

Burton-Jones, A. 2009. "Minimizing Method Bias through Programmatic Research," *MIS Quarterly* (33:3), pp. 445-471.

Butler, B. S., and Gray, P. H. 2006. "Reliability, Mindfulness, and Information Systems," *MIS Quarterly* (30:2), pp. 211-224.

Butterfield, K. D., Trevin, L. K., and Weaver, G. R. 2000. "Moral Awareness in Business Organizations: Influences of Issue-Related and Social Context Factors," *Human Relations* (53:7), pp. 981-1018.

Calders, T., and Verwer, S. 2010. "Three Naive Bayes Approaches for Discrimination-Free Classification," *Data Mining and Knowledge Discovery* (21:2), pp. 277-292.

Calders, T., and Žliobaitė, I. 2013. "Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures," in *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases*, B. Custers, T. Calders, B. Schermer and T. Zarsky (eds.). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 43-57.

Canadian Department of Justice. 2017. "Canadian Human Rights Act, 1985."

Carlson, D. S., Kacmar, K. M., and Wadsworth, L. L. 2002. "The Impact of Moral Intensity Dimensions on Ethical Decision Making: Assessing the Relevance of Orientation," *Journal of Managerial Issues* (14:1), pp. 15-30.

Chen, J., and Shen, X.-L. 2015. "Consumers' Decisions in Social Commerce Context: An Empirical Investigation," *Decision Support Systems* (79:1), pp. 55-64.

Chin, W. W. 1998. "The Partial Least Squares Approach to Structural Equation Modeling," in *Modern Methods for Business Research*, G.A. Marcoulide (ed.). New York, Ny: Lawrence Erlbaum Associates, pp. 295-336.

Chin, W. W. 2010. "How to Write up and Report Pls Analyses," in *Handbook of Partial Least Squares: Concepts, Methods and Application*, V. Esposito Vinzi, W.W. Chin and J. Henseler, Wang, H. (eds.). Berlin: Springer, pp. 655-690.

Chin, W. W., Marcolin, B. L., and Newsted, P. R. 2003. "A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study," *Information systems research* (14:2), pp. 189-217.

Chin, W. W., Thatcher, J. B., and Wright, R. T. 2012. "Assessing Common Method Bias: Problems with the Ulmc Technique," *MIS Quarterly* (36:3), pp. 1003-1019.

Chouldechova, A. 2017. "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments," *Big data* (5:2), pp. 153-163.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, J. R., Pant, L. W., and Sharp, D. J. 2001. "An Examination of Differences in Ethical Decision-Making between Canadian Business Students and Accounting Professionals," *Journal of Business Ethics* (30:4), pp. 319-336.

Collins, H. 2003. "Discrimination, Equality and Social Inclusion," *The modern law review* (66:1), pp. 16-43.

Côrte-Real, N., Oliveira, T., and Ruivo, P. 2017. "Assessing Business Value of Big Data Analytics in European Firms," *Journal of Business Research* (70:1), pp. 379-390.

Craft, J. L. 2013. "A Review of the Empirical Ethical Decision-Making Literature: 2004–2011," *Journal of Business Ethics* (117:2), pp. 221-259.

Crawford, K. 2013. "The Hidden Biases in Big Data." *Harvard Business Review* Retrieved 16/12/2015, from https://hbr.org/2013/04/the-hidden-biases-in-big-data

Cronbach, L. J. 1951. "Coefficient Alpha and the Internal Structure of Tests," *psychometrika* (16:3), pp. 297-334.

Cummings, M. L. 2006. "Automation and Accountability in Decision Support System Interface Design," *The journal of technology Studies* (32:1), pp. 23-31.

Custers, B. 2013. "Data Dilemmas in the Information Society: Introduction and Overview," in *Discrimination and Privacy in the Information Society*, B. Custers, T. Calders, B. Schermer and T. Zarsky (eds.). Berlin Heidelberg: Springer-Verlag, pp. 3-26.

Custers, B. H. M. 2004. *The Power of Knowledge: Ethical, Legal, and Technological Aspects of Data Mining and Group Profiling in Epidemiology*. Nijmegen, Netherlands: Wolf Legal Publishers.

Daniel, W. W. 1990. *Applied Nonparametric Statistics*, (2nd ed.). Boston, MA: PWS-Kent Publishing Company.

Danna, A., and Gandy, O. H. 2002. "All That Glitters Is Not Gold: Digging beneath the Surface of Data Mining," *Journal of Business Ethics* (40:4), pp. 373-386.

Datta, A., Tschantz, M. C., and Datta, A. 2015. "Automated Experiments on Ad Privacy Settings," *Proceedings on Privacy Enhancing Technologies* (2015:1), pp. 92-112.

DeLoache, J. S., Pierroutsakos, S. L., and Uttal, D. H. 2003. "The Origins of Pictorial Competence," *Current Directions in Psychological Science* (12:4), pp. 114-118.

DeLoache, J. S., Pierroutsakos, S. L., Uttal, D. H., Rosengren, K. S., and Gottlieb, A. 1998. "Grasping the Nature of Pictures," *Psychological Science* (9:3), pp. 205-210.

Diamantopoulos, A., and Siguaw, J. A. 2006. "Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration," *British Journal of Management* (17:4), pp. 263-282.

Dopplick, R. 2017. "New Statement on Algorithmic Transparency and Accountability by Acm U.S. Public Policy Council." Retrieved 21/4/2018, from https://techpolicy.acm.org/2017/01/new-statement-on-algorithmic-transparency-and-accountability-by-acm-u-s-public-policy-council/

Douglas, P. C., Davidson, R. A., and Schwartz, B. N. 2001. "The Effect of Organizational Culture and Ethical Orientation on Accountants' Ethical Judgments," *Journal of Business Ethics* (34:2), pp. 101-121.

Dutton, J. E., Dukerich, J. M., and Harquail, C. V. 1994. "Organizational Images and Member Identification," *Administrative science quarterly* (39:2), pp. 239-263.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. 2012. "Fairness through Awareness," *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, Cambridge, Massachusetts: ACM, pp. 214-226.

Ebrahimi, S., Ghasemaghaei, M., and Hassanein, K. 2016. "Understanding the Role of Data Analytics in Driving Discriminatory Managerial Decisions," *37th International Conference on Information Systems*, Dublin, Ireland.

Equal Employment Opportunity Coordinating Council, E. E. O. C. 1976. "Uniform Guidelines on Employee Selection Procedures," *Federal Register* (4:136), pp. 29016-29022.

European Union Legislation, 2011. "(a) European Convention on Human Rights, 1950; (B) Racial Equality Directive, 2000; (C) Employment Equality Directive, 2000; (D) Gender Goods and Services Directive, 2004; (E) Gender Employment Directive, 2006; (F) Equal Treatment Directive (Proposal), 2008.." http://eur-lex.europa.eu.

Falk, R. F., and Miller, N. B. 1992. *A Primer for Soft Modeling*. Akron, OH: University of Akron Press.

Federal Trade Commission. 2016. "Big Data: A Tool for Inclusion or Exclusion." Retrieved 25/4/2016

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., and Venkatasubramanian, S. 2015. "Certifying and Removing Disparate Impact," *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, NSW, Australia: ACM, pp. 259-268.

Ferrell, O. C., and Gresham, L. G. 1985. "A Contingency Framework for Understanding Ethical Decision Making in Marketing," *The Journal of Marketing* (49:3), pp. 87-96.

Ferrell, O. C., Gresham, L. G., and Fraedrich, J. 1989. "A Synthesis of Ethical Decision Models for Marketing," *Journal of Macromarketing* (9:2), pp. 55-64.

Fiske, S. T. 2000. "Stereotyping, Prejudice, and Discrimination at the Seam between the Centuries: Evolution, Culture, Mind, and Brain," *European Journal of Social Psychology* (30:3), pp. 299-322.

Flannery, B. L., and May, D. R. 2000. "Environmental Ethical Decision Making in the Us Metal-Finishing Industry," *Academy of Management journal* (43:4), pp. 642-662.

Fleischman, G., and Valentine, S. 2003. "Professionals' Tax Liability Assessments and Ethical Evaluations in an Equitable Relief Innocent Spouse Case," *Journal of Business Ethics* (42:1), pp. 27-44.

Fornell, C., and Larcker, D. F. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of marketing research* (18:February), pp. 39-50.

Frey, B. F. 2000. "The Impact of Moral Intensity on Decision Making in a Business Context," *Journal of Business Ethics* (26:3), pp. 181-195.

Gangadharan, S. P., Virginia, E., and Baroca, S. 2014. "Data and Discrimination: Collected Essays." Seattle, Washington: Open technology Institute, New America Foundation.

Gardner, J. 1989. "Liberals and Unlawful Discrimination," *Oxford J. Legal Stud.* (9:1), pp. 1-22.

Gates, S. W., Perry, V. G., and Zorn, P. M. 2002. "Automated Underwriting in Mortgage Lending: Good News for the Underserved?," *Housing Policy Debate* (13:2), pp. 369-391.

Gefen, D., and Straub, D. 2005. "A Practical Guide to Factorial Validity Using Pls-Graph: Tutorial and Annotated Example," *Communications of the Association for Information systems* (16:1), pp. 91-109.

Gefen, D., Straub, D., and Boudreau, M.-C. 2000. "Structural Equation Modeling and Regression: Guidelines for Research Practice," *Communications of the association for information systems* (4:7), pp. 1-70.

Ghasemaghaei, M., Ebrahimi, S., and Hassanein, K. 2018. "Data Analytics Competency for Improving Firm Decision Making Performance," *The Journal of Strategic Information Systems* (27:1), pp. 101-113.

Gioia, D. A., and Thomas, J. B. 1996. "Identity, Image, and Issue Interpretation: Sensemaking During Strategic Change in Academia," *Administrative science quarterly* (41:3), pp. 370-403.

Gregor, S., and Benbasat, I. 1999. "Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice," *MIS quarterly* (23:4), pp. 497-530.

Haines, R., Street, M. D., and Haines, D. 2008. "The Influence of Perceived Importance of an Ethical Issue on Moral Judgment, Moral Obligation, and Moral Intent," *Journal of Business Ethics* (81:2), pp. 387-399.

Hair, J., Hollingsworth, C. L., Randolph, A. B., and Chong, A. Y. L. 2017. "An Updated and Expanded Assessment of Pls-Sem in Information Systems Research," *Industrial Management & Data Systems* (117:3), pp. 442-458.

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. 2010. *Multivariate Data Analysis*, (7th ed.). Prentice hall Upper Saddle River, NJ.

Hanh, T. N. 1976. *The Miracle of Mindfulness*. Boston, MA: Beacon Press.

Hart, D., Atkins, R., and Ford, D. 1998. "Urban America as a Context for the Development of Moral Identity in Adolescence," *Journal of social issues* (54:3), pp. 513-530.

Haslam, N. 2006. "Dehumanization: An Integrative Review," *Personality and social psychology review* (10:3), pp. 252-264.

Hayes, S. C., Strosahl, K. D., and Wilson, K. G. 1999. *Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change*. New York, NY: Guilford Press.

Henderson, M. D., Fujita, K., Trope, Y., and Liberman, N. 2006. "Transcending the" Here": The Effect of Spatial Distance on Social Judgment," *Journal of personality and social psychology* (91:5), p. 845856.

Hong, Y.-Y., Benet-Martinez, V., Chiu, C.-Y., and Morris, M. W. 2003. "Boundaries of Cultural Influence: Construct Activation as a Mechanism for Cultural Differences in Social Perception," *Journal of Cross-Cultural Psychology* (34:4), pp. 453-464.

Hong, Y.-y., Morris, M. W., Chiu, C.-y., and Benet-Martinez, V. 2000. "Multicultural Minds: A Dynamic Constructivist Approach to Culture and Cognition," *American psychologist* (55:7), pp. 709-720.

Hunt, B. 2005. "Redlining," in: *The Electronic Encyclopedia of Chicago*. Chicago Historical Society.

Hunt, S. D., and Vitell, S. 1986. "A General Theory of Marketing Ethics," *Journal of macromarketing* (6:1), pp. 5-16.

Hunter, R. C., and Shoben, E. W. 1998. "Disparate Impact Discrimination: American Oddity or Internationally Accepted Concept?," *Berkeley Journal of Employment and Labor Law* (19:1), pp. 108-152.

Johnson, J. A. 2014. "The Ethics of Big Data in Higher Education," *International Review of Information Ethics* (21:1), pp. 3-10.

Jones, T. M. 1991. "Ethical Decision Making by Individuals in Organizations: An Issue-Contingent Model," *Academy of management review* (16:2), pp. 366-395.

Kabat-Zinn, J. 1982. "An Outpatient Program in Behavioral Medicine for Chronic Pain Patients Based on the Practice of Mindfulness Meditation: Theoretical Considerations and Preliminary Results," *General hospital psychiatry* (4:1), pp. 33-47.

Kabat-Zinn, J. 1994. *Wherever You Go, There You Are: Mindfulness Meditation in Everyday Life*. New York: NY: Hyperion.

Kabat-Zinn, J., and Hanh, T. N. 1990. *Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness*. New york:NY: Delacourte.

Kamiran, F., Calders, T., and Pechenizkiy, M. 2010. "Discrimination Aware Decision Tree Learning," *10th International Conference on Data Mining*, Sydney, NSW: IEEE, pp. 869-874.

Kamiran, F., Calders, T., and Pechenizkiy, M. 2013. "Techniques for Discrimination-Free Predictive Models," in *Discrimination and Privacy in the Information Society*. Berlin Heidelberg: Springer-Verlag, pp. 223-239.

Karacaer, S., Gohar, R., Aygün, M., and Sayin, C. 2009. "Effects of Personal Values on Auditor's Ethical Decisions: A Comparison of Pakistani and Turkish Professional Auditors," *Journal of Business Ethics* (88:1), pp. 53-64.

Kline, P. 2000. *Handbook of Psychological Testing*, (2nd ed.). London: Routledge.

Krieger, N. 1999. "Embodying Inequality: A Review of Concepts, Measures, and Methods for Studying Health Consequences of Discrimination," *International journal of health services* (29:2), pp. 295-352.

Krosnick, J. A., and Petty, R. E. 1995. "Attitude Strength: An Overview," in *Attitude Strength: Antecedents and Consequences*, R.E. Petty and J.A. Krosnick (eds.). Mahwah, NJ: Erlbaum, pp. 1-24.

Lehmann, D. R. 1988. "An Alternative Procedure for Assessing Convergent and Discriminant Validity," *Applied Psychological Measurement* (12:4), pp. 411-423.

Leitsch, D. L. 2004. "Differences in the Perceptions of Moral Intensity in the Moral Decision Process: An Empirical Examination of Accounting Students," *Journal of Business Ethics* (53:3), pp. 313-323.

Leitsch, D. L. 2006. "Using Dimensions of Moral Intensity to Predict Ethical Decision-Making in Accounting," *Accounting Education: an international journal* (15:2), pp. 135-149.

Liang, H., Saraf, N., Hu, Q., and Xue, Y. 2007. "Assimilation of Enterprise Systems: The Effect of Institutional Pressures and the Mediating Role of Top Management," *MIS quarterly* (31:1), pp. 59-87.

Liberman, N., Sagristano, M. D., and Trope, Y. 2002. "The Effect of Temporal Distance on Level of Mental Construal," *Journal of experimental social psychology* (38:6), pp. 523-534.

Liberman, N., Trope, Y., and Stephan, E. 2007. "Psychological Distance," in *Social Psychology: Handbook of Basic Principles*, A.W. Kruglanski and E.T. Higgins (eds.). New York:NY: Guilford Press, pp. 353-381.

Linehan, M. 1993. *Cognitive-Behavioral Treatment of Borderline Personality Disorder*. New York, NY: Guilford press.

Loe, T. W., Ferrell, L., and Mansfield, P. 2000. "A Review of Empirical Studies Assessing Ethical Decision Making in Business," *Journal of business ethics* (25:3), pp. 185-204.

Loe, T. W., and Weeks, W. A. 2000. "An Experimental Investigation of Efforts to Improve Sales Students' Moral Reasoning," *Journal of Personal Selling & Sales Management* (20:4), pp. 243-251.

Lyon, D. 2003. *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination*. London: Routledge.

Maier, C., Laumer, S., Weinert, C., and Weitzel, T. 2015. "The Effects of Technostress and Switching Stress on Discontinued Use of Social Networking Services: A Study of Facebook Use," *Information Systems Journal* (25:3), pp. 275-308.

Makkonen, T. 2002. "Multiple, Compound and Intersectional Discrimination: Bringing the Experiences of the Most Marginalized to the Fore," 1102, Institute for Human Rights, A˚ bo Akademi University.

Martin, K. E. 2015. "Ethical Issues in the Big Data Industry," *MIS Quarterly Executive* (14:2), pp. 67-85.

Massey, D. S., and Denton, N. A. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

May, D. R., and Pauli, K. P. 2002. "The Role of Moral Intensity in Ethical Decision Making a Review and Investigation of Moral Recognition, Evaluation, and Intention," *Business & Society* (41:1), pp. 84-117.

Mayer-Schönberger, V., and Cukier, K. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, NY: Houghton Mifflin Harcourt.

McAfee, A., and Brynjolfsson, E. 2012. "Big Data: The Management Revolution," *Harvard business review* (90:10), pp. 60-66.

McDonald, C. 2016. "Is There a Sexist Data Crisis?," in: *BBC News*. United Kingdom.

Menard, S. 1995. *Applied Logistic Regression Analysis*. Thousand Oaks, CA: Sage.

Mencl, J., and May, D. R. 2009. "The Effects of Proximity and Empathy on Ethical Decision-Making: An Exploratory Investigation," *Journal of Business Ethics* (85:2), pp. 201-226.

Meyers, L. S., Gamst, G., and Guarino, A. J. 2006. *Applied Multivariate Research: Design and Interpretation*. Thousand Oaks, CA: Sage publications.

Milgram, S. 1974. *Obedience to Authority*. New York: Harper & Row.

Moberg, D., and Caldwell, D. F. 2007. "An Exploratory Investigation of the Effect of Ethical Culture in Activating Moral Imagination," *Journal of Business Ethics* (73:2), pp. 193-204.

Morris, S. A., and McDonald, R. A. 1995. "The Role of Moral Intensity in Moral Judgments: An Empirical Investigation," *Journal of Business Ethics* (14:9), pp. 715-726.

Newell, S., and Marabelli, M. 2014. "The Crowd and Sensors Era: Opportunities and Challenges for Individuals, Organizations, Society, and Researchers," *35th International Conference on Information Systems*, Auckland, NZ.

Newell, S., and Marabelli, M. 2015. "Strategic Opportunities (and Challenges) of Algorithmic Decision-Making: A Call for Action on the Long-Term Societal Effects of 'Datification'," *The Journal of Strategic Information Systems* (24:1), pp. 3-14.

Nisbett, R. E., and Ross, L. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Nunnally, J. C. 1978. *Psychometric Theory*. New York: McGraw-Hill.

Nunnally, J. C., and Bernstein, I. H. 1994. *Psychometric Theory*. Ney York: McGraw-Hill.

Nussbaum, S., Liberman, N., and Trope, Y. 2006. "Predicting the near and Distant Future," *Journal of Experimental Psychology: General* (135:2), pp. 152-161.

O'Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY: Crown Publishing Group.

O'Fallon, M. J., and Butterfield, K. D. 2005. "A Review of the Empirical Ethical Decision-Making Literature: 1996–2003," *Journal of business ethics* (59:4), pp. 375-413.

Osofsky, M. J., Bandura, A., and Zimbardo, P. G. 2005. "The Role of Moral Disengagement in the Execution Process," *Law and Human Behavior* (29:4), pp. 371-393.

Paolillo, J. G., and Vitell, S. J. 2002. "An Empirical Investigation of the Influence of Selected Personal, Organizational and Moral Intensity Factors on Ethical Decision Making," *Journal of Business Ethics* (35:1), pp. 65-74.

Paulhus, D. L. 1984. "Two-Component Models of Socially Desirable Responding," *Journal of personality and social psychology* (46:3), pp. 598-609.

Paulhus, D. L. 1991. "Measurement and Control of Response Bias," in *Measures of Personality and Social Psychological Attitudes,* J.P. Robinson, P.R. Shaver and L.S. Wrightsman (eds.). New York: Academic Press, pp. 17-59.

Pedhazur, E. J., and Schmelkin, L. P. 1991. *Measurement, Design, and Analysis: An Integrated Approach*, (1st ed.). New York: Lawrence Erlbaum Associates.

Pedreschi, D., Ruggieri, S., and Turini, F. 2009. "Measuring Discrimination in Socially-Sensitive Decision Records," *Proceedings of the 2009 SIAM International Conference on Data Mining*, Sparks, Nevada: Society for Industrial and Applied Mathematics (SIAM), pp. 581-592.

Pedreschi, D., Ruggieri, S., and Turini, F. 2013. "The Discovery of Discrimination," in *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases,* B. Custers, T. Calders, B. Schermer and T. Zarsky (eds.). Berlin Heidelberg: Springer-Verlag, pp. 91-108.

Pedreshi, D., Ruggieri, S., and Turini, F. 2008. "Discrimination-Aware Data Mining," *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*: ACM, pp. 560-568.

Petter, S., Straub, D., and Rai, A. 2007. "Specifying Formative Constructs in Information Systems Research," *MIS Quarterly* (31:4), pp. 623-656.

Petty, R., and Wegener, D. 1998. "Attitude Change: Multiple Roles for Persuasion Variables," in *The Handbook of Social Psychology*, D. Gilbert, S. Fiske and G. Lindzey (eds.). New York: McGraw-Hill, pp. 323–390.

Petty, R. E., and Cacioppo, J. T. 1986a. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer-Verlag.

Petty, R. E., and Cacioppo, J. T. 1986b. "The Elaboration Likelihood Model of Persuasion," *Advances in experimental social psychology* (19:1), pp. 123-205.

Petty, R. E., Haugtvedt, C. P., and Smith, S. M. 1995. "Elaboration as a Determinant of Attitude Strength: Creating Attitudes That Are Persistent, Resistant, and Predictive of Behavior," in *Attitude Strength: Antecedents and Consequences*, R.E. Petty and J.A. Krosnick (eds.). Mahwah, NJ: Erlbaum.

Podesta, J., Pritzker, P., Moniz, E. J., Holdren, J., and Zients, Z. 2014. "Big Data: Seizing Opportunities, Preserving Values," Executive Office of the President, Washington, DC.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., and Podsakoff, N. P. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of applied psychology* (88:5), pp. 879-903.

Podsakoff, P. M., and Organ, D. W. 1986. "Reports in Organizational Research: Problems and Prospects," *Journal of Management studies* (27), pp. 305-327.

Randall, D. M., and Fernandes, M. F. 1991. "The Social Desirability Response Bias in Ethics Research," *Journal of Business Ethics* (10:11), pp. 805-817.

Rest, J. R. 1986. *Moral Development: Advances in Research and Theory*. New York: Praeger.

Reynolds, S. J. 2006. "Moral Awareness and Ethical Predispositions: Investigating the Role of Individual Differences in the Recognition of Moral Issues," *Journal of Applied Psychology* (91:1), pp. 233-243.

Roldán, J. L., and Sánchez-Franco, M. J. 2012. "Variance-Based Structural Equation Modeling: Guidelines for Using Partial Least Squares," in *Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems*, M. Mora, O. Gelman, A. Steenkamp and M. Raisinghani (eds.). Hershey, PA: Information Science Reference, pp. 193-221.

Romei, A., and Ruggieri, S. 2014. "A Multidisciplinary Survey on Discrimination Analysis," *Knowledge Engineering Review* (29:5), pp. 582–638.

Romei, A., Ruggieri, S., and Turini, F. 2013. "Discrimination Discovery in Scientific Project Evaluation: A Case Study," *Expert Systems with Applications* (40:15), pp. 6064-6079.

Rottig, D., Koufteros, X., and Umphress, E. 2011. "Formal Infrastructure and Ethical Decision Making: An Empirical Investigation and Implications for Supply Management," *Decision Sciences* (42:1), pp. 163-204.

Ruedy, N., and Schweitzer, M. 2010. "In the Moment: The Effect of Mindfulness on Ethical Decision Making," *Journal of Business Ethics* (95:1), pp. 73-87.

Ruggieri, S., Pedreschi, D., and Turini, F. 2010. "Dcube: Discrimination Discovery in Databases," *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, Indianapolis, Indiana: ACM, pp. 1127-1130.

Russom, P. 2011. "Big Data Analytics," in: *TDWI Best Practices Report, Fourth Quarter*.

Ryan, R. M., and Deci, E. L. 2000. "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being," *American psychologist* (55:1), pp. 68-78.

Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. 2014. "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms," *Data and discrimination: converting critical concerns into productive inquiry*, Seattle, WA, pp. 1-23.

Schermer, B. W. 2011. "The Limits of Privacy in Automated Profiling and Data Mining," *Computer Law & Security Review* (27:1), pp. 45-52.

Schrage, M. 2014. "Big Data's Dangerous New Era of Discrimination." *Harvard Business Review* Retrieved 4/4/2016, from https://hbr.org/2014/01/big-datas-dangerous-new-era-of-discrimination/

Segal, Z. V., Williams, J., and Mark, G. 2002. *Mindfulness-Based Cognitive Therapy for Depression-a New Approach to Preventing Relapse*. New York, NY: Guilford.

Shang, W., Jiang, Z. M., Hemmati, H., Adams, B., Hassan, A. E., and Martin, P. 2013. "Assisting Developers of Big Data Analytics Applications When Deploying on Hadoop Clouds," *Proceedings of the 2013 International Conference on Software Engineering*, San Francisco, CA: IEEE Press, pp. 402-411.

Silver, M. S. 1991. "Decisional Guidance for Computer-Based Decision Support," *MIS Quarterly* (15:1), pp. 105-122.

Singer, M., Mitchell, S., and Turner, J. 1998. "Consideration of Moral Intensity in Ethicality Judgements: Its Relationship with Whistle-Blowing and Need-for-Cognition," *Journal of Business Ethics* (17:5), pp. 527-541.

Singhapakdi, A., Rao, C., and Vitell, S. J. 1996a. "Ethical Decision Making: An Investigation of Services Marketing Professionals," *Journal of Business Ethics* (15:6), pp. 635-644.

Singhapakdi, A., Vitell, S. J., and Franke, G. R. 1999. "Antecedents, Consequences, and Mediating Effects of Perceived Moral Intensity and Personal Moral Philosophies," *Journal of the Academy of Marketing Science* (27:1), pp. 19-36.

Singhapakdi, A., Vitell, S. J., and Kraft, K. L. 1996b. "Moral Intensity and Ethical Decision-Making of Marketing Professionals," *Journal of Business Research* (36:3), pp. 245-255.

Stauffer, J. M., and Buckley, M. R. 2005. "The Existence and Nature of Racial Bias in Supervisory Ratings," *Journal of Applied Psychology* (90:3), pp. 586-591.

Stenberg, G. 2006. "Conceptual and Perceptual Factors in the Picture Superiority Effect," *European Journal of Cognitive Psychology* (18:6), pp. 813-847.

Straub, D., Boudreau, M.-C., and Gefen, D. 2004. "Validation Guidelines for Is Positivist Research," *The Communications of the Association for Information Systems* (13:24), pp. 380-427.

Street, M. D., Douglas, S. C., Geiger, S. W., and Martinko, M. J. 2001. "The Impact of Cognitive Expenditure on the Ethical Decision-Making Process: The Cognitive Elaboration Model," *Organizational Behavior and Human Decision Processes* (86:2), pp. 256-277.

Sweeney, B., Arnold, D., and Pierce, B. 2010. "The Impact of Perceived Ethical Culture of the Firm and Demographic Variables on Auditors' Ethical Evaluation and Intention to Act Decisions," *Journal of Business Ethics* (93:4), pp. 531-551.

Tabachnick, B. G., and Fidell, L. S. 1996. *Using Multivariate Statistics*, (3rd ed.). New York: HarperCollins.

Tene, O., and Polonetsky, J. 2013. "Judged by the Tin Man: Individual Rights in the Age of Big Data," *11 Journal of Telecommunications and High Technology Law* (11), pp. 351-368.

Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., and Lauro, C. 2005. "Pls Path Modeling," *Computational statistics & data analysis* (48:1), pp. 159-205.

Trevino, L. K. 1986. "Ethical Decision Making in Organizations: A Person-Situation Interactionist Model," *Academy of management Review* (11:3), pp. 601-617.

Trevino, L. K., and Brown, M. E. 2004. "Managing to Be Ethical: Debunking Five Business Ethics Myths," *The Academy of Management Executive* (18:2), pp. 69-83.

Treviño, L. K., Butterfield, K. D., and McCabe, D. L. 1998. "The Ethical Context in Organizations: Influences on Employee Attitudes and Behaviors," *Business Ethics Quarterly* (8:3), pp. 447-476.

Trope, Y., and Liberman, N. 2003. "Temporal Construal," *Psychological review* (110:3), pp. 403-421.

Trope, Y., and Liberman, N. 2010. "Construal-Level Theory of Psychological Distance," *Psychological review* (117:2), pp. 440-463.

U.K. Legislation. 2011. "(a) Sex Discrimination Act, 1975, (B) Race Relation Act, 1976." http://www.legislation.gov.uk/.

Urbach, N., and Ahlemann, F. 2010. "Structural Equation Modeling in Information Systems Research Using Partial Least Squares," *JITTA: Journal of Information Technology Theory and Application* (11:2), p. 5.

Valentine, S. R., and Bateman, C. R. 2011. "The Impact of Ethical Ideologies, Moral Intensity, and Social Context on Sales-Based Ethical Reasoning," *Journal of Business Ethics* (102:1), pp. 155-168.

Veale, M., and Binns, R. 2017. "Fairer Machine Learning in the Real World: Mitigating Discrimination without Collecting Sensitive Data," *Big Data & Society* (4:2), pp. 1-17.

Venkatesh, V., and Agarwal, R. 2006. "Turning Visitors into Customers: A Usability-Centric Perspective on Purchase Behavior in Electronic Channels," *Management Science* (52:3), pp. 367-382.

Wakslak, C. J., Trope, Y., Liberman, N., and Alony, R. 2006. "Seeing the Forest When Entry Is Unlikely: Probability and the Mental Representation of Events," *Journal of Experimental Psychology: General* (135:4), pp. 641-653.

Wang, W.-T., Wang, Y.-S., and Liu, E.-R. 2016. "The Stickiness Intention of Group-Buying Websites: The Integration of the Commitment–Trust Theory and E-Commerce Success Model," *Information & Management* (53:5), pp. 625-642.

Wang, W., and Benbasat, I. 2007. "Recommendation Agents for Electronic Commerce: Effects of Explanation Facilities on Trusting Beliefs," *Journal of Management Information Systems* (23:4), pp. 217-246.

Watley, L. D., and May, D. R. 2004. "Enhancing Moral Intensity: The Roles of Personal and Consequential Information in Ethical Decision-Making," *Journal of Business Ethics* (50:2), pp. 105-126.

Weaver, K. M., and Ferrell, O. C. 1977. "The Impact of Corporate Policy in Reported Ethical Beliefs and Behavior of Marketing Practitioners," *American Marketing Association (AMA)*, Atlanta, GA, pp. 477–481.

Westen, D. 1999. *Psychology: Mind, Brain, and Culture*, (2nd ed.). New York:NY: John Wiley & Sons.

Wetzels, M., Odekerken-Schröder, G., and Van Oppen, C. 2009. "Using Pls Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Empirical Illustration," *MIS quarterly*), pp. 177-195.

Williams, B. A., Brooks, C. F., and Shmargad, Y. 2018. "How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications," *Journal of Information Policy* (8), pp. 78-115.

Williams, L. J., Edwards, J. R., and Vandenberg, R. J. 2003. "Recent Advances in Causal Modeling Methods for Organizational and Management Research," *Journal of management* (29:6), pp. 903-936.

Woods, D. 2011. "Linkedin's Monica Rogati on "What Is a Data Scientist?"." Retrieved 10/4/2018, from https://www.forbes.com/sites/danwoods/2011/11/27/linkedins-monica-rogati-on-what-is-a-data-scientist/#2c19d8f83e15

Wotruba, T. R. 1990. "A Comprehensive Framework for the Analysis of Ethical Behavior, with a Focus on Sales Organizations," *Journal of Personal Selling & Sales Management* (10:2), pp. 29-42.

Yan, D., Sengupta, J., and Hong, J. 2016. "Why Does Psychological Distance Influence Construal Level? The Role of Processing Mode," *Journal of Consumer Research* (43:4), pp. 598-613.

Zhang, J., Chiu, R., and Wei, L. 2009. "Decision-Making Process of Internal Whistleblowing Behavior in China: Empirical Evidence and Implications," *Journal of Business Ethics* (88:1), pp. 25-41.

Zliobaite, I. 2015. "Measuring Discrimination in Algorithmic Decision Making," *Data Mining and Knowledge Discovery* (31:4), pp. 1060-1089.

Žliobaitė, I., and Custers, B. 2016. "Using Sensitive Personal Data May Be Necessary for Avoiding Discrimination in Data-Driven Decision Models," *Artificial Intelligence and Law* (24:2), pp. 183-201.

APPENDIX A. Survey Questions

Acceptance of the Recommendation

1. Please indicate whether you accept the tool's recommendation about the recommended group of employees to be sent to a one-week training abroad on leadership skills?

☐ Accept

☐ Reject

Screening Questions

2. Which of the following best describes your role in your organization?

☐ Upper management

☐ Middle management

☐ Junior management

☐ Administrative staff

☐ Support staff

☐ Skilled labor

☐ Consultant

☐ Temporary employee

☐ Researcher

☐ Student

☐ Self-employed

☐ Other …………………

3. Please indicate the number of employees in your organization who directly or indirectly report to you.

☐ 0 employees

☐ 1 employee

☐ 2 to 3 employees

☐ 4 to 5 employees

☐ 6 to 9 employees

☐ 10 to 49 employees

☐ 50 to 99 employees

☐ 100 to 499 employees

☐ 500 to 999 employees

☐ 1000 or more employees

Please indicate the degree to which you agree or disagree with the following statements regarding your experience with using the data analytics tool you just used. When answering these questions, please keep in mind that there are no right or wrong answers, so please answer the questions as honestly as possible.

Please note that the phrase "decision" refers to the decision you, with the support of the data analytics tool, made today regarding the selection of the employees to be sent abroad to a one-week training program on Effective Leadership in a Sales Organization.

Reference	Factor	Adapted Measurement Instruments	7-Point Likert Scale Anchors						
			Strongly disagree	**Disagree**	**Somewhat disagree**	**Neither agree nor disagree**	**Somewhat agree**	**Agree**	**Strongly agree**
Barnett (2001)	Perceived proximity	The employees depicted in the context in which you used the DA tool today are: Similar to me	○	○	○	○	○	○	○
		Alike me	○	○	○	○	○	○	○
		Different from me	○	○	○	○	○	○	○
Reynolds (2006, Study 2)	Recognition of the moral issue	There were very important ethical aspects to the decision I was asked to make today	○	○	○	○	○	○	○
		The decision I was asked to make today clearly did NOT involve ethics or moral issues	○	○	○	○	○	○	○
		The decision I was asked to make could be described as a moral issue	○	○	○	○	○	○	○

Reference	Factor	Adapted Measurement Instruments	7-Point Likert Scale Anchors						
			Strongly disagree	**Disagree**	**Somewhat disagree**	**Neither agree nor disagree**	**Somewhat agree**	**Agree**	**Strongly agree**
Brown and Ryan (2003)	Mindfulness	I could be experiencing some emotion and not be conscious of it until sometime later.	○	○	○	○	○	○	○
		I break or spill things because of carelessness, not paying attention, or thinking of something else.	○	○	○	○	○	○	○
		I find it difficult to stay focused on what’s happening in the present.	○	○	○	○	○	○	○
		I tend to walk quickly to get where I’m going without paying attention to what I experience along the way.	○	○	○	○	○	○	○
		I tend not to notice feelings of physical tension or discomfort until they really grab my attention.	○	○	○	○	○	○	○

Reference	Factor	Adapted Measurement Instruments	7-Point Likert Scale Anchors						
			Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Brown and Ryan (2003)	Mindfulness	I forget a person's name almost as soon as I've been told it for the first time.	○	○	○	○	○	○	○
		It seems I am ''running on automatic'' without much awareness of what I'm doing.	○	○	○	○	○	○	○
		I rush through activities without being really attentive to them.	○	○	○	○	○	○	○
		I get so focused on the goal I want to achieve that I lose touch with what I am doing right now to get there.	○	○	○	○	○	○	○
		I do jobs or tasks automatically, without being aware of what I'm doing.	○	○	○	○	○	○	○
		I find myself listening to someone with one ear, doing something else at the same time.	○	○	○	○	○	○	○

Reference	Factor	Adapted Measurement Instruments	7-Point Likert Scale Anchors						
			Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Brown and Ryan (2003)	Mindfulness	I drive to places on ‘‘automatic pilot’’ and then wonder why I went there.	○	○	○	○	○	○	○
		I find myself preoccupied with the future or the past.	○	○	○	○	○	○	○
		I find myself doing things without paying attention.	○	○	○	○	○	○	○
		I snack without being aware that I’m eating.	○	○	○	○	○	○	○
(Treviño et al. 1998)	Ethical culture	Management in our organization disciplines unethical behavior when it occurs.	○	○	○	○	○	○	○
		Employees in our organization perceive that people who violate the ethics code of the organization still get formal organizational rewards.	○	○	○	○	○	○	○

Reference	Factor	Adapted Measurement Instruments	7-Point Likert Scale Anchors						
			Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
(Treviño et al. 1998)	Ethical culture	Penalties for unethical behavior are strictly enforced in our organization.	○	○	○	○	○	○	○
		Unethical behavior is punished in our organization.	○	○	○	○	○	○	○
		Top managers of our organization portray high ethical standards.	○	○	○	○	○	○	○
		People of integrity are rewarded in our organization.	○	○	○	○	○	○	○
		The ethics code serves as "window dressing" only in our organization.	○	○	○	○	○	○	○
		Top managers of our organization regularly show that they care about ethics.	○	○	○	○	○	○	○

Reference	Factor	Adapted Measurement Instruments	7-Point Likert Scale Anchors						
			Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
(Treviño et al. 1998)	Ethical culture	Top managers of our organization are models of unethical behavior.	○	○	○	○	○	○	○
		Ethical behavior is the norm in our organization.	○	○	○	○	○	○	○
		Top managers of our organization guide decision making in an ethical direction.	○	○	○	○	○	○	○
		The ethics code serves only to maintain the organization's public image.	○	○	○	○	○	○	○
		Ethical behavior is rewarded in our organization.	○	○	○	○	○	○	○
		Ethics code requirements are consistent with informal organizational norms.	○	○	○	○	○	○	○

Reference	Factor	Adapted Measurement Instruments	7-Point Likert Scale Anchors						
			Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Paulhus (1991)	Impression Management *(Control Variable)*	I sometimes tell lies if I have to.	○	○	○	○	○	○	○
		I never cover up my mistakes.	○	○	○	○	○	○	○
		There have been occasions when I have taken advantage of someone.	○	○	○	○	○	○	○
		I always obey laws, even if I'm unlikely to get caught.	○	○	○	○	○	○	○
		I have done things that I don't tell other people about.	○	○	○	○	○	○	○
(Aquino and Reed 2002)	Moral Identity (Internalization) *(Control Variable)*	It would make me feel good to be a person who has these characteristics.	○	○	○	○	○	○	○

Reference	**Factor**	**Adapted Measurement Instruments**	**7-Point Likert Scale Anchors**						
			Strongly disagree	**Disagree**	**Somewhat disagree**	**Neither agree nor disagree**	**Somewhat agree**	**Agree**	**Strongly agree**
(Aquino and Reed 2002)	Moral Identity (Internalization) *(Control Variable)*	Being someone who has these characteristics is an important part of who I am.	○	○	○	○	○	○	○
		I would be ashamed to be a person who had these characteristics.	○	○	○	○	○	○	○
		Having these characteristics is not really important to me.	○	○	○	○	○	○	○
		I strongly desire to have these characteristics.	○	○	○	○	○	○	○
(Aquino and Reed 2002)	Moral Identity (Symbolization) *(Control Variable)*	I often wear clothes that identify me as having these characteristics.	○	○	○	○	○	○	○
		The types of things I do in my spare time (e.g., hobbies, volunteer activities) clearly identify me as having these characteristics.	○	○	○	○	○	○	○

Reference	Factor	Adapted Measurement Instruments	7-Point Likert Scale Anchors						
			Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
(Aquino and Reed 2002)	Moral Identity (Symbolization) *(Control Variable)*	The kinds of books and magazines that I read identify me as having these characteristics.	○	○	○	○	○	○	○
		The fact that I have these characteristics is communicated to others by my membership in certain organizations.	○	○	○	○	○	○	○
		I am actively involved in activities that communicate to others that I have these characteristics.	○	○	○	○	○	○	○
Manipulation Check	Demographic Transparency	I felt that I had demographic transparency regarding the subjects of my decision.	○	○	○	○	○	○	○
		I felt that I had information about the demographics of the subjects of my decision.	○	○	○	○	○	○	○
		I was able to demographically visualize the subjects of my decision.	○	○	○	○	○	○	○

APPENDIX B- Consent Form

LETTER OF INFORMATION / CONSENT

A Study of/about: Data Analytics Tools Design

Investigators:

Student Investigator:	Faculty Supervisor:
Sepideh Ebrahimi	Dr. Khaled Hassanein
DeGroote School of Business	DeGroote School of Business
McMaster University	McMaster University
Hamilton, Ontario, Canada	Hamilton, Ontario, Canada
(905) 525-9140 ext. 26048	**(905) 525-9140 ext. 23956**
E-mail: s.ebrahimi@mcmaster.ca	E-mail: hassank@mcmaster.ca

Purpose of the Study: We are conducting this study as a part of a Ph.D. dissertation that aims to find out how Data Analytics Users interact with Data Analytic tools. This research will result in guidelines for design of data analytics tools.

Procedures involved in the Research: This study will last approximately 30 minutes. If you volunteer to participate in this study, you will be asked to use a data analytic tool to decide about selecting a group of employees to be sent abroad for a training program on leadership skills. Subsequently, you will be asked to complete an online survey. In the survey, you will be asked to respond to closed-ended questions about your experience of using the data analytic tool to make the decision as well as your decision-making style and your current working environment. After completing the survey, you will be asked to respond to open-ended questions to gather basic background information about your experience.

Potential Harms, Risks or Discomforts: The risks involved in participating in this study are minimal. There are no foreseeable physical, psychological, emotional, social, or financial risks associated with this study. Please note that you do not need to answer questions that you do not want to answer or that make you feel uncomfortable. You can also stop taking part in the study (withdraw) at any time.

Potential Benefits: By participating in this study, you will help to discover ways to improve data analytics tools and their recommendations. Particularly, the results of this study will help researchers and practitioners understand if several data analytics design factors impact data analytics users' perceptions and evaluation of the system.

Compensation: You will be compensated by Research Now as outlined in Research Now's compensation policy. You must complete the survey before you can enter your e-mail address into the sweepstakes. Please note that you are still eligible for compensation if you elect not to answer some of the questions in the survey. See https://www.researchnow.com/terms-and-conditions/ for further information about the compensation process.

Confidentiality: The survey is anonymous. All information collected from you will be kept secure and in strict confidence. Only the researchers named above will have access to the data, which will be stored securely. Participants will not be identified individually in any reports or analyses resulting from this study.

Participation and Withdrawal: Your participation in this study is voluntary. If you decide to be part of the study, you can withdraw at any time. If you decide to withdraw, there will be no consequences to you and none of your survey responses will be collected or stored. You will not be eligible for compensation, if you decide to withdraw.

Information about the Study Results: We expect to have this study completed by approximately summer 2018. The results of the study will be posted on the MacSphere website (McMaster University Libraries Institutional Repository): https://macsphere.mcmaster.ca/

Questions about the Study

If you have questions or need more information about the study itself, please contact us at: s.ebrahimi@mcmaster.ca or (905) 525-9140 ext. 26048.

This study has been reviewed by the McMaster University Research Ethics Board and received ethics clearance.

If you have concerns or questions about your rights as a participant or about the way the study is conducted, please contact:

McMaster Research Ethics Secretariat

Telephone: (905) 525-9140 ext. 23142

c/o Research Office for Administrative Development and Support

E-mail: ethicsoffice@mcmaster.ca

CONSENT

I understand the information provided for the study "Data Analytics Design" as described herein. My questions have been answered to my satisfaction, and by clicking on the " Yes" button below, I agree to participate in this study. I understand that if I agree to participate in this study, I may withdraw from the study at any time.

"I agree to participate."

"I do not agree to participate."

APPENDIX C- Instructions

Purpose of the Study: We are conducting this study as a part of a Ph.D. dissertation that aims to better understand how data analytics users' interact with data analytics tools. This research will result in guidelines for design of such tools.

Data Analytics (DA) is the process of examining large data sets in order to uncover patterns, associations, and other useful information to help organizations make more informed business decisions.

This study includes a data analytics tool designed to be used for Human Resource purposes to help organizations with internal hiring and/or promotions as well as selecting employees for various purposes (e.g., training programs). More specifically, the tool is designed for use by managers who are interested in identifying the employees who should be considered for specific new assignments within the organization requiring a specific set of qualifications. Being able to identify such an initial list of employees for managers' consideration enables them to be efficient and effective in their human resources decisions.

The data analytics tool initially uses historical data to identify the criteria (e.g., education level, active years of service, etc.) that are most associated with success in different types of situations (e.g., positions, trainings). In so doing, the data analytics tool devises a statistical algorithm for each type of situation based on a weighted combination of its relevant critera. This algorithm could then be used from thereon to identify the most suited employees to consider for a particular situation.

The DA tool is subsequently used as follows:

- When a manager uses the DA tool to identify employees to considser for a particular situation, the appropriate type of algorithm matching the situation type will be selected. Then the tool applies the algorithm on all the employees in the organization's data base who currently work in a relevant position type.

- The tool then returns the list of all the employees it considered along with a recommended decision as to whether or not each employee should be considered for the situation in question (e.g., position, promotion, training program).
- The tool also includes functionality to allow its users to sort and filter the results based on different criteria (e.g., years of service, etc.).

It is important to note that the algorithms developed by the DA tool for these purposes are statistical in nature and rely on historical and current data. As such, they are relatively accurate but not perfect.

Therefore, in this study, we are interested in testing the above-described tool. You are asked to use this tool, to find employees who should be considered for a one-week training program on leadership skills. The goal is to choose a group of 20 employees to be sent abroad for the training. Attending the training will bring about great experience for the selected employees. In addition, they will be more likely to receive promotions in the future.

After using the tool, you will be asked to decide whether you accept the recommended list generated by the tool. You will also be asked to complete an online survey regarding your expereicne with using the tool.

Next

APPENDIX D- Debriefing Letter

Demographic Transparency to Combat Data Analytics Discriminatory Recommendations

Investigators:

Student Investigator:	**Faculty Supervisor:**
Sepideh Ebrahimi	Dr. Khaled Hassanein
DeGroote School of Business	DeGroote School of Business
McMaster University	McMaster University
Hamilton, Ontario, Canada	Hamilton, Ontario, Canada
(905) 525-9140 ext. 26048	**(905) 525-9140 ext. 23956**
E-mail:	E-mail:
s.ebrahimi@mcmaster.ca	hassank@mcmaster.ca

Thank you for taking this survey. Your time and effort are much appreciated. Your answers are a valuable part of this research.

This study seeks to investigate the potential discriminatory recommendations generated by Data Analytics tools and whether the users of these tools will approve such recommendations. Recently Data Analytics has been blamed for contributing to discriminatory decision making in organizations by several researchers, practitioners, and even governments. However, much of such discriminatory decisions are made unknowingly based on accepting potentially discriminatory recommendations that are made because of biased or non-representative data sets, and inadvertent modeling procedures. In order to ameliorate this issue, this study investigates the impact of the availability of aggregated demographic information to the users of such tools on decreasing the likelihood of their acceptance of potential discriminatory data analytics recommendations. We expect that users who receive aggregated demographic information about the subjects of their decisions will be less likely to approve the potentially discriminatory recommendation. In addition, since receiving photos of the subject of their decision can help users feel less distant from them, we expect that users who receive such photos will be less likely to approve the discriminatory recommendation.

In order to properly test our hypothesis, we could not provide you with all of these details prior to your participation in the experiment. This ensures that your responses to the questions in this study were not influenced by prior knowledge about the full purpose of the study. If we had told you the full purposes of our study, your decision about whether or not to approve the recommendation you received from the data analytics tool could have been affected. We hope you understand.

Please note that although the purpose of this study was not fully disclosed to you, everything else on the consent form is correct. This includes the ways in which we will keep your data confidential. "This survey is anonymous. All information collected from you will be kept secure and in strict confidence. Only the researchers named above will have access to the data, which will be stored securely. Participants will not be identified individually in any reports or analyses resulting from this study"

Please note that the data set and tool you used in this study are totally fictitious and therefore, no one has been advantaged or disadvantaged in the real world as a result of your decision. In addition, your participation in this study is completely anonymous as no identifiable information has been collected.

If you have any questions or concerns about this study, its purpose or procedures, please feel free to contact the researchers, Sepideh Ebrahimi (s.ebrahimi@mcmaster.ca) and/or Dr. Khaled Hassanein (hassank@mcmaster.ca).

THANK YOU AGAIN FOR YOUR PARTICIPATION.

POST-DEBRIEFING CONSENT

I have been debriefed about the research project entitled "Demographic Transparency to Combat Data Analytics Discriminatory Recommendations" and I have had an opportunity to read the debriefing information provided. I agree to allow the data collected during my participation in this research project to be used, understanding that I am doing so voluntarily and that confidentiality and anonymity will be maintained.

Yes "I DO want my data to be included in this study."

No "I DO NOT want my data to be included in this study."