Understanding the Role of Data Analytics in Driving Discriminatory Managerial Decisions

Research-in-Progress

Sepideh Ebrahimi
McMaster University
Hamilton, Canada
s.ebrahimi@mcmaster.ca

Maryam Ghasemaghaei
McMaster University
Hamilton, Canada
ghasemm@mcmaster.ca

Khaled Hassanein
McMaster University
Hamilton, Canada
hassank@mcmaster.ca

Abstract

Data analytics has been accused of contributing to discriminatory managerial decisions in organizations' marketing strategies. To date, most studies have focused on the technical antecedents of such discriminations and, therefore, little is known about the role of human factors in making these discriminatory decisions. This work-in-progress study aims at addressing this gap by opening the black box between data analytics use in organizations and making discriminatory decisions. Drawing upon the theory of moral disengagement, we posit that four dimensions of moral disengagement, namely, dehumanization, euphemistic labeling, displacement of responsibility, and disregard of consequences are the mechanisms through which the use of data analytics tools in organizations could bring about discriminatory decisions. Moreover, data size and employees' competency are discussed as having moderating impacts on some of these mechanisms. A survey-based methodology to empirically validate the proposed model is outlined. Potential contributions to theory and practice are delineated.

Keywords: Data analytics, Discrimination, Unethical decision making, Dehumanization, Euphemistic labeling, Displacement of responsibility, Disregard of consequences

Introduction

The last decade has witnessed an ever-increasing adoption of e-commerce by organizations and a widespread diffusion of digitized devices among customers, which have enabled the organizations to collect an increasing amount of data about their current and potential customers. This data is then used by organizations to make 'data-driven' decisions, which as McAfee and Brynjolfsson (2012) argue, is superior to traditional 'HIPPO' (highest-paid person's opinion) style decision making. More specifically, the data is being used to target and personalize products and services for customers based on the patterns found, and recommendations made by data analytics tools. Data analytics is the process of analyzing large amounts of data using computer systems to discover hidden 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). While analyzing customers' data and personalizing products and services based on data-driven insights can provide businesses with strategic opportunities, there have been societal concerns raised about such methods of classifying customers for the purpose of providing personalized products and services (Newell and Marabelli 2015).

One such societal concern is discrimination, which refers to "the process by which a member, or members, of a socially defined group is, or are, treated differently (especially unfairly) because of his/her/their membership of that group" (Krieger 1999, p. 301). Discrimination is considered a highly unethical behavior and its adverse consequences in personal, social, and organizational contexts have been studied extensively. Nonetheless, in recent years, use of data analytics tools to make algorithmic decisions has been accused of exacerbating discrimination in societies (Danna and Gandy 2002; Johnson 2014; Lyon 2003; Newell and Marabelli 2015). While the use of analytics tools had raised concerns about making discriminatory decisions, the rise of big data has compounded the issue. For example, in the pre-big data era, decision makers, could figure out if (due to housing segregation) neighborhood is a good proxy for race. However, using data with the high volume and variety provided by big data and advanced data analytics tools, it is possible that some combination of "likes" on Facebook, network of friends, and musical tastes can be used to predict membership in 'protected classes' (e.g. race, gender, age) and companies can use these to discriminate in their decision making (Barocas and Selbst 2016).

One main method used by data analytics tools from which discrimination originates is profiling (Danna and Gandy 2002; Newell and Marabelli 2014). A number of researchers, therefore, have strived to develop methodologies (for classification, generating decision trees, etc.) to decrease the level of discrimination that data analytics tools bring about in their results (e.g. Calders and Verwer 2010; Kamiran et al. 2010; Pedreshi et al. 2008). The suggested improved methodologies, though effective in reducing the technical issues that lead to generating discriminatory results by data analytics tools, are not sufficient to ensure that the use of such tools to make data-driven decisions will indeed lead to making ethical and non-discriminatory decisions. The evidence for such insufficiency is manifested in the recent scholarly and practitioners' publications raising awareness about the potential of making discriminatory decisions using data analytics tools (Federal Trade Commission 2016; Newell and Marabelli 2015; Schrage 2014).

In organizations, it is ultimately the managers and decision makers' responsibility to make sure that their data-driven decisions are free of discrimination. Consequently, we believe that one way to advance the literature on the discrimination stemming from use of data analytics tools is to open the black box of mediating mechanisms that intervene between using data analytics tools and making discriminatory decisions. Identifying these key mediating factors can be an important step toward designing interventions geared toward reducing the incidence of discriminatory decisions in the context of data-driven decision making using data analytic tools.

The important role of cognitions in driving individual behavior is generally well-recognized (Gollwitzer and Bargh 1996). Therefore, identifying the cognitions from which unethical discriminatory decision making can potentially stem is a good starting point for understanding why such decisions are made in organizations. To this end, we employ the theory of moral disengagement, which elaborates why individuals engage in an unethical and/or socially unacceptable behavior although its immorality is widely admitted (Bandura 1986, 1990). The theory of moral disengagement is, therefore, a proper theoretical lens for understanding why certain managers engage in the unethical behavior of making discriminatory decisions while using data analytics tools. This theory has been used to study a variety of unethical behaviors including but not limited to engaging in undermining at work (Lee et al. 2016), decisions to commit fraud and self-serving decisions in the workplace (Moore et al. 2012), deception (Barsky 2011), and bullying (Hymel et al. 2005).

Drawing on this theory, we posit that four cognitive mechanisms, namely, dehumanization, euphemistic labeling, displacement of responsibility, and disregard of consequences are associated with using data analytics tools. These mechanisms, as suggested by the theory of moral disengagement and as shown by several researches, will facilitate making unethical discriminatory decisions. Next, we discuss the theoretical background of this study and then turn to the theoretical underpinning of the proposed research model. Methodology in support of collecting data and pertinent analyses is presented in the fourth section. Concluding remarks close the paper.

Theoretical Background

The notion of moral disengagement was developed by Bandura (1999) as an extension of social cognitive theory. Social cognitive theory provides an agentic view of human behavior whereby individuals control their thoughts and behaviors through self-regulatory mechanisms (Bandura 1986). According to this theory, most people have developed personal standards of moral behavior, which governs a system of self-monitoring of and self-reaction to one's conduct. In other words, individuals use their personal moral standards to monitor and judge their own behavior and as a result are guided toward decent behavior and deterred from misbehavior. However, moral self-regulation functions only if it is activated. Bandura (1999) argued that the self-regulatory mechanisms can be selectively turned on and off and suggested that moral disengagement is the main switch for the deactivation process.

The notion of moral disengagement explains the reason for individuals being able to engage in inappropriate conduct. As suggested by cognitive dissonance theory (Festinger 1957), when an individual engages in a behavior that is not consistent with their beliefs, a state of psychological tension is induced. The person then seeks to reduce the dissonance between the belief and the behavior in order to ameliorate the uncomfortable feeling. Moral disengagement is used as a tool to align an individual's beliefs with their behavior. Through moral disengagement, individuals are freed from self-sanctions as well as the compunction that would follow when behavior violates internal standards (Detert et al. 2008).

"As long as self-sanctions override the force of external inducements, behavior is kept in line with personal standards" (Bandura 1990, p. 28). However, when the external inducement is strong (e.g. higher profit, personal promotion, higher performance), such conflicts are resolved through selective disengagement of self-sanctions. Researchers have long strived to describe and categorize the techniques used for such disengagement of self-sanctions. For instance, Sykes and Matza (1957) discuss five techniques of *neutralization* (i.e. denial of responsibility, denial of injury, denial of victim, condemnation of the condemners, and appeal to higher loyalties). Ashforth and Anand (2003) add three other techniques (i.e. legality, metaphor of the ledger, and refocusing attention) to make a list of eight such techniques. Bandura in his theory of moral disengagement suggests that people override their self-restraints against engagement in detrimental behavior in order to perform unethical, antisocial, and self-serving behaviors through eight mechanisms for *moral disengagement*: moral justification, euphemistic labeling, advantageous comparison, displacement of responsibility, diffusion of responsibility, disregarding or distorting the consequences, dehumanization, and attribution of blame (Bandura 1986). The mechanisms suggested by various researchers, despite their differences, overlap to a great extent. In this article we suggest that while making unethical discriminatory decisions using data analytics tools, four mechanisms of moral disengagement play an important role. Specifically, we propose that use of data analytics tools is associated with increases in dehumanization, euphemistic labeling, displacement of responsibility, and disregard of consequences. These four mechanisms are particularly relevant to the use of data analytics tools for decision making due to the contextual conditions facilitated by such tools. We define and discuss each of these four moral disengagement mechanisms and discuss our arguments for them being possibly activated by using data analytics tools in the next section.

Research Model and Hypotheses Development

Figure 1 presents the proposed research model. As the figure depicts, individual cognitions in the form of moral disengagement are suggested to play a mediating role in linking data analytics use and discriminatory decision making. Specifically, this study argues that the use of data analytics tools can lead to making discriminatory decisions in organizations through four mechanisms of moral disengagement (i.e. dehumanization, euphemistic labeling, displacement of responsibility, and disregard of consequences). Since the positive association of moral disengagement on unethical behavior and decision making (e.g., discrimination) has been discussed extensively in the literature (e.g. Barsky 2011; Detert et al. 2008; Moore 2008; Osofsky et al. 2005), we do not include it in our hypotheses. Next, we elaborate on the logic of our proposed research model.

Dehumanization

Dehumanization represents “the denial of qualities associated with meaning, interest, and compassion” toward others (Barnard 2001, p. 98). Individuals, who engage in dehumanizing others, do not perceive the human qualities of others and therefore do not view them as persons with feelings, hopes, and concerns but as “subhuman objects” (Bandura 1999). Therefore, dehumanization nullifies self-restraints that operate through feelings of empathy and compassion (Osofsky et al. 2005). Several studies have shown that dehumanization is associated with an increase in the likelihood to engage in a number of unethical and antisocial behavior such as social loafing (Alnuaimi et al. 2010), aggression (Bandura et al. 1975; Rudman and Mescher 2012), and proclivity to torture prisoners of war (Viki et al. 2013). The reason for the connection between dehumanization and such unethical behavior lies in the fact that dehumanization and the psychological distancing that it creates make it easier to deny the impact of one’s unethical behavior on the victims (Ashforth and Anand 2003).

In a comprehensive review of the literature, Haslam (2006) argues that there are two main types of dehumanization: (i) denying uniquely human characteristics (e.g. civility, rationality, and maturity); or (ii) denying human nature characteristic (e.g. individuality, emotional responsiveness, and interpersonal warmth). Haslam refers to the former as animalistic dehumanization and the latter as mechanistic dehumanization and argues that while animalistic dehumanization implies disgust, mechanistic dehumanization is associated with indifference. In this article, we suggest that use of data analytics tools increases the likelihood of engaging in mechanistic dehumanization, which in turn is associated with unethical decision making (e.g. discriminatory decision making).

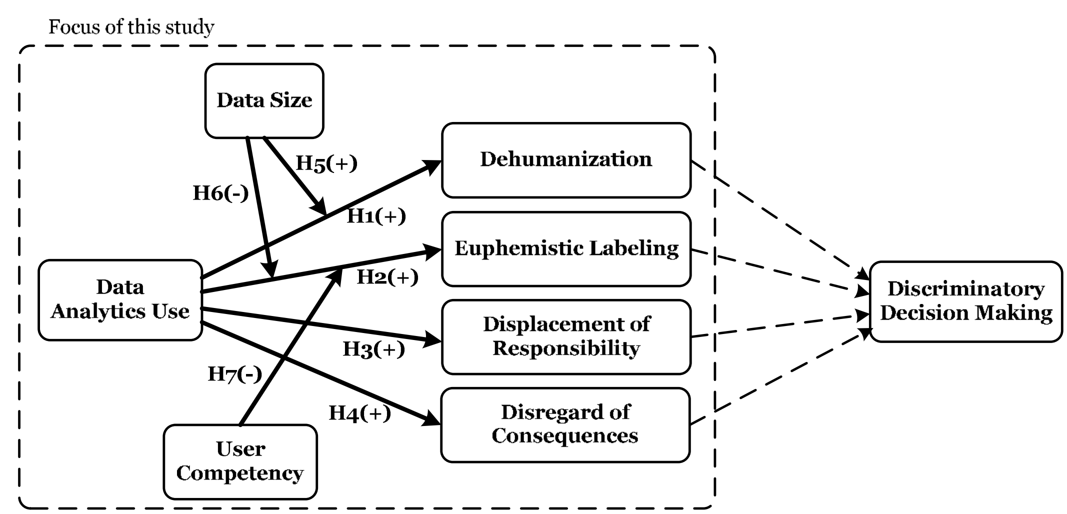

Figure 1. Research Model

Refusing to ascribe an identity to a person is a pivotal element of dehumanization (Ashforth and Anand 2003; Haslam 2006; Kelman 1976). Dehumanization is, therefore, salient in the context of technology and especially in the context of using data analytic tools for decision making. Data analytics tools tend to treat individuals as a set of records, each with a number of attributes, which are incapable of carrying individuals' individuality. They represent a "generalized other" (Mead 1934). Data analytics tools, thus, facilitate dehumanization through representing individuals as a set of data and therefore, using data analytics is expected to positively influence dehumanization. When individuals psychologically dehumanize others (e.g. customers, job applicants), they are less likely to feel remorse for making unethical discriminatory decisions that deprive the dehumanized others from having access to products, promotions, etc. Therefore, we posit that:

Hypothesis 1: *Use of data analytics tools for decision making will increase the likelihood of dehumanization.*

Euphemistic Labeling

Euphemistic labeling is another mechanism for moral disengagement. Euphemistic labeling involves cognitive restructuring of reprehensible behavior in a way that increases its moral acceptability (Bandura 1986). Using morally neutral language, individuals make their otherwise unethical behaviors seem less harmful or even benign (Detert et al. 2008). For instance, "strategic misrepresentation" is used to euphemize lying to business competitors (Safire 1979) and "team players" is used as a euphemistic term for those who collude with unethical actions within corrupt organizations (Jackall 1988). Euphemistic labeling is, therefore, a key self-deceptive method that allows individuals to behave unethically in organizations (Tenbrunsel and Messick 2004).

Use of data analytics tools for decision making involves euphemistic labeling in two main ways. First, many of the decisions made as a result of drawing on algorithms are discriminatory although they are referred to as personalization (Danna and Gandy 2002). For example, imagine a case that Schrage (2014) discusses where data analytics tools in a company show that single Asian, Hispanic, and African-American women with urban postal codes are most likely to complain to a company about the quality of the products and services. In addition, Asian and Hispanic complainers who end up being satisfied with the resolution of their complaints tend to be among the most profitable customers but African-American women do not. In that case, a decision to provide Asian, and Hispanic females with preferential treatment over African-American women when handling complaints, will be an unethical and discriminatory decision.

Second, although some organizations might be reluctant to make such discriminatory decisions based on gender, race, etc. using data analytics tools can lead to "indirect discrimination" (Pedreshi et al. 2008). For instance, the decision to deny credit to residents of a specific neighborhood asking for a car loan entails discrimination if being a resident in that neighborhood is strongly correlated with some discriminatory conditions, such as being a member of an ethnic minority (Pedreshi et al. 2008). The same goes for the process of hiring an employee from a pool of applicants. A company might decide to include applicants' living addresses in their hiring algorithm, if the company's historical records show that employees who live closer to their jobs tend to stay at their jobs longer than those who live farther away. Such a decision may lead to racial discrimination particularly since different neighborhoods can have different racial compositions. In the two aforementioned cases, the decision maker may not acknowledge making discriminatory decisions and may call it classification that is geared toward protecting the organization and safeguarding its resources. In light of the above discussion, we suggest that:

Hypothesis 2: *Use of data analytics tools for decision making will increase the likelihood of euphemistic labeling.*

Displacement of Responsibility

Moral control operates more strongly when people admit their role and their detrimental actions in causing harm to others. Displacement of responsibility functions by minimizing the agentive role in the harm that one engenders (Bandura 1999). With displacement of responsibility, individuals view their actions as springing from social pressures and dictates of others rather than being their personal responsibility (Bandura 2001). As a result, individuals' self-censure is reduced because they are no longer actual agents of their actions and the responsibility for those harmful actions are shifted onto someone else (Hinrichs et al. 2012). In other words, the unethical behavior is ascribed to compelling circumstances and, thus, not viewed as a personal decision (Rogers 2001).

Displacement of responsibility can play a major role in making unethical decisions using computer systems. Cummings (2006) introduces the notion of "moral buffer", which is related but not identical to moral disengagement. She argues that "A moral buffer adds an additional layer of ambiguity and possible diminishment of accountability and responsibility [to moral disengagement] through an artifact or process, such as a computer interface or automated recommendations" (Cummings 2006, p. 26). One of the boosters of moral buffer is when people assign moral agency to computers in spite of the fact that they are inanimate objects. In fact, individuals generally have a tendency to anthropomorphize computers and perceive some human properties in technological artifacts (Dryer 1999; Reeves and Nass 1996). As a result, when problems occur, computers can be seen as at least partially responsible (Friedman and Millett 1997).

The fact that computer systems can diminish users' sense of their personal moral agency and responsibility may lead to erosion of accountability (Friedman and Kahn 1997). As a result, individuals might tend not to hold themselves accountable for the consequences of their computer use (Friedman and Kahn 1997). An example of displacement of responsibility while using computer systems has been reported in regard to clinical decision support systems. Acute Physiology and Chronic Health Evaluation (APACHE) system is one such system that is employed to determine the stage of an illness where a person should be removed from life support systems as further treatments would be futile (Helft et al. 2000). Cummings (2006) argues that using such systems can lead to individuals distance themselves from a very difficult decision and shift the moral responsibility of their decisions to the computer artifact.

We argue that the same situation is true in regard to using data analytics tools in organizations for decision making. Users of such systems may follow the recommendations made by data analytics tools and attribute the responsibility of any unethical discriminatory decisions they make to the tools. As a result, we hypothesize that:

Hypothesis 3: *Use of data analytics tools for decision making will increase the likelihood of displacement of responsibility.*

Disregard of consequences

When people pursue activities that are harmful to others, they are likely to avoid facing the harm they cause. If individuals disregard the negative results of their unethical conduct, they are less likely to recognize an inconsistency between their moral system and their conduct. For example, one may know that fraud is immoral in a purely rational sense, but through moral disengagement might be capable of redefining fraud as a justified act of personal gain with negligible consequences (Stevens et al. 2012). As another example imagine a customer that may tell themselves that no one will be harmed if they do not report an error in their favor because "this little bit of money doesn't affect anything in a huge company like X" (Detert et al. 2008, p. 376). In short, as long as the harmful results of one's behavior are disregarded, the self-censure is unlikely to be activated as the detrimental consequences of the harmful conduct are disbelieved (Bandura 1990; Bandura et al. 1996).

The role of physical distance in increasing the disregard of the harmful consequences has been acknowledged by several scholars. 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". Jones (1991) in a seminal paper on the issue-contingent model of ethical decision making defined proximity as the feeling of social, cultural, psychological, or physical nearness that an individual has for victims of the evil act in question and asserted that perceived proximity is an influential factor in avoiding to engage in an unethical behavior. In a series of obedience experiments, Milgram (1974) empirically shows that unethical behavior is inversely related to committing unethical behavior. Milgram's subjects ("teachers") were ordered to administer powerful shocks to learners when they failed to answer certain questions correctly. Milgram found that increased physical proximity of the teachers and learners significantly reduced the occurrence of unethical behavior (i.e. giving the learner electric shocks) from 62.5 percent to 30 percent.

Computers and technological artifacts play a pivotal role in disregarding the consequences of an unethical behavior. Computer systems enable individuals to remotely perform mass destruction in wars without seeing and hearing the suffering they cause (Bandura 1999). "Some technological devices provide the remote distance that makes it easier to kill. These devices can be TV and video screens, thermal sights, or some other mechanical apparatus that provides a psychological buffer" (Cummings 2006, p. 27). We argue that use of data analytics tools can play a similar, though less harsh, role in organizational decision making. Use of such tools leads to distancing the decision maker and the victims of unethical, discriminatory decisions and as a result the suffering of those who are discriminated against can be masked and therefore invisible to the decision maker. For example, imagine the case where a decision maker in an insurance organization, using data analytics tools, classifies customers based on their characteristics into different groups and, therefore, decides to charge young male customers higher than young female customers for similar insurance coverage due to young males being more risk-prone in general. In that case, the decision maker by pushing some buttons on a computer screen makes a discriminatory decision against young male customers; however he/she is not present to see how the average male customer feels when he figures out that he has been discriminated against[1]. Therefore, we posit that:

Hypothesis 4: *Use of data analytics tools for decision making will increase the likelihood of disregard of consequences.*

Data Size and User Competency

In the context of this study, data size refers to the size of the population (e.g. number of customers) as well as the number of attributes (e.g. gender, age, address) about which, data have been collected. We argue that an increase in data volume strengthens the relationship between the use of data analytics tools and dehumanization. The adverse impact of population size on increasing dehumanization has already been discussed in the literature on various topics including but not limited to social loafing in team work (Alnuaimi et al. 2010), and violating human rights (Poe et al. 2006). Such an increase in dehumanization is mainly due to the fact that as the size of a population increases, one has difficulty ascribing an identity to each of the members of the population, which itself is a pivotal element in dehumanization (Haslam 2006; Kelman 1976).

It might be argued that even with a small data set, the user of a data analytics tool is unlikely to ascribe a distinguishable identity to a member of the population of interest (e.g. a customer). However, we suggest that as the data size increases, it is more likely that the user views the population as a set of records with a number of attributes that should be classified and/or clustered for the purpose of making further decisions. Therefore, we hypothesize that:

Hypothesis 5: *Data Size positively moderates the impact of use of data analytics tools for decision making on dehumanization.*

Increases in data volume, on the other hand, can lead to lowering the likelihood of euphemistic labeling. A higher number of attributes in the data set enables the user to base their decisions on individuals' attributes other than the ones that will lead to a discriminatory decision. In that case it is less likely that the decision maker needs to camouflage their discriminatory decisions by labeling them euphemistically. Furthermore, by having more attributes, the data analytics tools can come up with a higher number of decision alternatives. Having a number of various alternative is a required factor for ethical decision making (Hunt and Vitell 1986). For example, if only customers' gender, age, country of residence, and email address are known, almost any classification that the software produces would be discriminatory. On the other hand, if the data analytics tool is to classify customers when their website browsing history, purchase history, etc. are known, it is more likely that the classification made is based on attributes other than customers' gender and age. As a result, we argue that:

Hypothesis 6: *Data Size negatively moderates the impact of use of data analytics tools for decision making on euphemistic labeling.*

---

[1] It is noteworthy that since December 2012 insurers in EU are no longer allowed to use statistical evidence about gender differences to set premiums (Newell and Marabelli 2015).

Data analytics users' competency refers to the level of users' knowledge of the nature of the data set and expertise in working with data analytics tools. Having higher data analytics competency can mitigate the adverse effect of data analytics tools on euphemistic labeling. When a user has higher expertise in using the tool, they are more likely to develop various alternatives for decision making, which in turn can lead to a lower need for euphemistic labeling of their discriminatory decisions. Furthermore, when a user has a higher level of knowledge of the data set and tool expertise, they are more likely to figure out hidden relationships in the data, which in turn may result in making decisions that might not look discriminatory but are in fact discriminatory in nature. For example, they might find out that customers' residential address is related to their ethnicity and, therefore, a classification based on customers' home address may bring about racial discrimination.

Hypothesis 7: *Users' competency negatively moderates the impact of use of data analytics tools for decision making on euphemistic labeling.*

Methodology

This study will utilize a cross-sectional survey of middle-managers who use data analytics tools to make marketing decisions. This sampling choice was made since the concept of discriminatory decision making has been most discussed in relation to marketing decision making. We will develop a data collection instrument, where the constructs in the proposed research model will be operationalized using measures adapted from validated instruments. The four dimensions of the moral engagement included in this study will be measured using the instrument developed by Bandura et al. (1996), which will be modified to fit the context of this study. Data analytics use will be measured using a 2-item scale adapted from Venkatesh et al. (2008). We will measure data size using a single-item scale and user's competency adopting the 9-item scale from Marcolin et al. (2000). A pilot survey will be used to test and refine the measurement instrument. In addition, ethics approval from the ethics research board at the authors' university will be secured prior to data collection.

It is noteworthy that we will strive 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 1991). 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 under-report undesirable behaviors" (Paulhus 1991, p. 4). In order to minimize the potential social desirability effect, we will assure respondents that the survey is anonymous and confidential. Furthermore, we will follow Paulhus (1991) suggestion to use the Impression Management subscale of the Balanced Inventory of Desirable Responding (BIDR) to test for social desirability effects. We chose the impression management scale to be consistent with other studies (Flannery and May 2000; Treviño et al. 1998; Watley and May 2004) and also to reduce respondents' fatigue.

In order to validate the instrument, we will examine construct reliability, convergent validity, and discriminant validity for all constructs. In addition, we will examine the common method bias using Herman's single-factor test, as per Podsakoff et al. (2003) because all measures will be collected at one point in time. To test the research model, we will use Partial Least Squares (PLS), a structural equation modeling (SEM) technique as it is more suited for exploratory research (Gefen et al. 2000). In addition, we will conduct a saturated model analysis to figure out if there are any possible significant relationships not hypothesized in the model (Chin et al. 2003).

According to Roldán and Sánchez-Franco (2012), to have a sufficient statistical power of 0.80 to detect a medium effect size (f=.25), 74 samples are required for this study. Further, when PLS is used for data analysis, the sample size must be at least ten times the number of items used to measure the construct with the highest number of items in the research model (Gefen et al. 2000). In this study, users' competence has the highest number of items (i.e. nine items) in the research model. Therefore, the minimum sample size for this study must be 90. As there will be some potential outliers and to account for spoiled or incomplete responses, 150 samples will be collected for this study.

Conclusion

Our goal in this research is to better understand the cognitive mechanisms that are activated by using data analytics tools. These mechanisms can deactivate moral self-regulation and allow individuals to make discriminatory decisions more easily. This study will make several contributions to both theory and practice. To the best of our knowledge, our research is the first to propose that moral disengagement mechanisms are facilitated by the use of data analytics tools for decision making in organizations. Thereby, we contribute to the social psychology literature by extending the theory of moral disengagement to a new context. In addition, this research contributes to the literature on unethical decision making in organizations by operationalizing the aforementioned four relevant mechanisms of moral disengagement (i.e. dehumanization, euphemistic labeling, displacement of responsibility, and disregard of consequences) in the context of organizational decision making. Furthermore, by opening the black box of the association between unethical discriminatory decisions making in organizations and use of data analytics tools, researchers and practitioners can move forward with designing interventions that target those mechanisms. These interventions can be in both technical (e.g. interface design) and organizational (e.g. goal definition, ethical climate, raising awareness). Last but not least, we identify two moderators that can impact the relationships between data analytics use and two of the moral disengagement mechanisms (i.e. dehumanization and euphemistic labeling). We suggest that companies need to pay attention to the data size as it can play the role of a double-edged sword. While data size can attenuate the positive association between data analytics use and euphemistic labeling, it may increase the likelihood of dehumanization. Therefore, in the presence of high volumes of data (which is the case in the big data era), companies should take more caution as to how the data size influences their employees' perception of their customers' individuality and humanization.

The result of this study along with psychological studies that look into measuring individuals' propensity to morally disengage can help practitioners who seek to lower the level of discriminatory decisions made by their data analytics users. This is due to the fact that moral disengagement has been discussed to have certain personal characteristics antecedents (e.g. empathy, locus of control, moral identity) (Detert et al. 2008) and, therefore, different individuals have various levels of propensity to morally disengage (Moore et al. 2012). It is also notable that moral disengagement is not a stable trait but rather is malleable to external influences over time (Paciello et al. 2008). Therefore, future research can look into how different interventions (e.g. training, goal definition) can reduce data analytics users' level of moral disengagement and therefore the likelihood of discriminatory decisions.

One potential limitation of this study is that we did not include ethical culture of the organization in our proposed model. It is generally argued that organizational factors (e.g. ethical culture) play a pivotal role in driving employees' (un)ethical behavior (Ashforth and Anand 2003; Gigerenzer 2010; Haidt 2001). Nonetheless, we intentionally excluded ethical culture from our model to avoid a priming effect, which would likely occur if the respondents are asked about their organization's ethical culture (e.g. the availability of codes of ethics). Our focus in this study is on the individual level of analysis and asking individuals about their perception of ethical culture in their organization will inevitably lead to the results being affected by a priming effect.

References

Alnuaimi, O. A., Robert, L. P., and Maruping, L. M. 2010. "Team Size, Dispersion, and Social Loafing in Technology-Supported Teams: A Perspective on the Theory of Moral Disengagement," *Journal of Management Information Systems* (27:1), pp. 203-230.

Ashforth, B. E., and Anand, V. 2003. "The Normalization of Corruption in Organizations," *Research in Organizational Behavior* (25), pp. 1-52.

Bandura, A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. 1990. "Selective Activation and Disengagement of Moral Control," *Journal of Social Issues* (46:1), pp. 27-46.

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

Bandura, A. 2001. "Social Cognitive Theory of Mass Communication," *Media psychology* (3:3), pp. 265-299.

Bandura, A., Barbaranelli, C., Caprara, G. V., and Pastorelli, C. 1996. "Mechanisms of Moral Disengagement in the Exercise of Moral Agency," *Journal of personality and social psychology* (71:2), pp. 364-374.

Bandura, A., Underwood, B., and Fromson, M. E. 1975. "Disinhibition of Aggression through Diffusion of Responsibility and Dehumanization of Victims," *Journal of research in personality* (9:4), pp. 253-269.

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.

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

Barsky, A. 2011. "Investigating the Effects of Moral Disengagement and Participation on Unethical Work Behavior," *Journal of business ethics* (104:1), pp. 59-75.

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

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.

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

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.

Detert, J. R., Treviño, L. K., and Sweitzer, V. L. 2008. "Moral Disengagement in Ethical Decision Making: A Study of Antecedents and Outcomes," *Journal of Applied Psychology* (93:2), pp. 374-391.

Dryer, D. C. 1999. "Getting Personal with Computers: How to Design Personalities for Agents," *Applied Artificial Intelligence* (13:3), pp. 273-295.

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

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Evanston, IL: Row and Peterson.

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.

Friedman, B., and Kahn, P. H. 1997. "Human Agency and Responsible Computing: Implications for Computer System Design," in *Human Values and the Design of Computer Technology,* B. Friedman (ed.). Stanford, CA: CSLI Publications, pp. 221-235.

Friedman, B., and Millett, L. I. 1997. "Reasoning About Computers as Moral Agents: A Research Note," in *Human Values and the Design of Computer Technology,* B. Friedman (ed.). Stanford, CA,: CSLI Publications, pp. 201-205.

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.

Gigerenzer, G. 2010. "Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality," *Topics in cognitive science* (2:3), pp. 528-554.

Gollwitzer, P. M., and Bargh, J. A. 1996. *The Psychology of Action: Linking Cognition and Motivation to Behavior*. New York: Guilford Press.

Haidt, J. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological review* (108:4), pp. 814-834.

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

Helft, P. R., Siegler, M., and Lantos, J. 2000. "The Rise and Fall of the Futility Movement," *New England Journal of Medicine* (343:4), pp. 293-296.

Hinrichs, K. T., Wang, L., Hinrichs, A. T., and Romero, E. J. 2012. "Moral Disengagement through Displacement of Responsibility: The Role of Leadership Beliefs," *Journal of Applied Social Psychology* (42:1), pp. 62-80.

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

Hymel, S., Rocke-Henderson, N., and Bonanno, R. A. 2005. "Moral Disengagement: A Framework for Understanding Bullying among Adolescents," *Journal of Social Sciences* (8:1), pp. 1-11.

Jackall, R. 1988. *Moral Mazes: The World of Corporate Managers*. New York: Oxford University Press.

Johnson, J. A. 2014. "The Ethics of Big Data in Higher Education," *International Review of Information Ethics* (7:1), 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.

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.

Kelman, H. G. 1976. "Violence without Moral Restraint: Reflections on the Dehumanization of Victims and Victimizers," in *Varieties of Psychohistory,* G.M. Kren and L.H. Rappoport (eds.). New York: Springer, pp. 282-314.

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.

Lee, K., Kim, E., Bhave, D. P., and Duffy, M. K. 2016. "Why Victims of Undermining at Work Become Perpetrators of Undermining: An Integrative Model," *Journal of Applied Psychology* (101:6), p. 915.

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

Marcolin, B. L., Compeau, D. R., Munro, M. C., and Huff, S. L. 2000. "Assessing User Competence: Conceptualization and Measurement," *Information Systems Research* (11:1), pp. 37-60.

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

Mead, G. H. 1934. *Mind, Self and Society*. Chicago, Ill: University of Chicago Press.

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

Moore, C. 2008. "Moral Disengagement in Processes of Organizational Corruption," *Journal of Business Ethics* (80:1), pp. 129-139.

Moore, C., Detert, J. R., Klebe Treviño, L., Baker, V. L., and Mayer, D. M. 2012. "Why Employees Do Bad Things: Moral Disengagement and Unethical Organizational Behavior," *Personnel Psychology* (65:1), pp. 1-48.

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.

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.

Paciello, M., Fida, R., Tramontano, C., Lupinetti, C., and Caprara, G. V. 2008. "Stability and Change of Moral Disengagement and Its Impact on Aggression and Violence in Late Adolescence," *Child Development* (79:5), pp. 1288-1309.

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.

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.

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.

Poe, S. C., Rost, N., and Carey, S. C. 2006. "Assessing Risk and Opportunity in Conflict Studies a Human Rights Analysis," *Journal of Conflict Resolution* (50:4), pp. 484-507.

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.

Reeves, B., and Nass, C. 1996. *How People Treat Computers, Television, and New Media Like Real People and Places*. Stanford, CA: CSLI Publications.

Rogers, M. K. 2001. "A Social Learning Theory and Moral Disengagement Analysis of Criminal Computer Behavior: An Exploratory Study," in: *Department of Psychology*. Canada: University of Manitoba.

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.

Rudman, L. A., and Mescher, K. 2012. "Of Animals and Objects Men's Implicit Dehumanization of Women and Likelihood of Sexual Aggression," *Personality and Social Psychology Bulletin* (38:6), pp. 734-746.

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

Safire, W. 1979. "The Fine Art of Euphemism," in: *San Francisco Chronicle*. p. 13.

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/

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.

Stevens, G. W., Deuling, J. K., and Armenakis, A. A. 2012. "Successful Psychopaths: Are They Unethical Decision-Makers and Why?," *Journal of Business Ethics* (105:2), pp. 139-149.

Sykes, G. M., and Matza, D. 1957. "Techniques of Neutralization: A Theory of Delinquency," *American sociological review* (22:6), pp. 664-670.

Tenbrunsel, A. E., and Messick, D. M. 2004. "Ethical Fading: The Role of Self-Deception in Unethical Behavior," *Social Justice Research* (17:2), pp. 223-236.

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.

Venkatesh, V., Brown, S. A., Maruping, L. M., and Bala, H. 2008. "Predicting Different Conceptualizations of System Use: The Competing Roles of Behavioral Intention, Facilitating Conditions, and Behavioral Expectation," *MIS quarterly* (32:3), pp. 483-502.

Viki, G. T., Osgood, D., and Phillips, S. 2013. "Dehumanization and Self-Reported Proclivity to Torture Prisoners of War," *Journal of Experimental Social Psychology* (49:3), pp. 325-328.

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.