Implicit Bias in Predictive Data Profiling Within Recruitments

Anders Persson

Implicit Bias in Predictive Data Profiling within Recruitments

Anders Persson

Division of Visual Information and Interaction, Department of Information Technology, Uppsala University, Sweden
anders.persson@it.uu.se

Abstract: Recruiters today are often using some kind of tool with data mining and profiling, as an initial screening for successful candidates. Their objective is often to become more objective and get away from human limitation, such as implicit biases versus underprivileged groups of people. In this explorative analysis there have been three potential problems identified, regarding the practice of using these predictive computer tools for hiring. First, that they might miss the best candidates, as the employed algorithms are tuned with limited and outdated data. Second, is the risk of directly or indirectly discriminate candidates, or, third, failure to give equal opportunities for all individuals. The problems are not new to us, and from this theoretical analysis and from other similar work; it seems that algorithms and predictive data mining tools have similar kinds of implicit biases as humans. Our human limitations, then, does not seem to be limited to us humans.

Keywords: Data mining · Social exclusion · Discrimination · Implicit bias · Recruitment · Big Data · People analytics · Machine-learning.

1 Introduction

How can a company looking for a specific person for a specific role find the most likely candidate to succeed among thousands of applicants? This is a problem facing many recruiters that are either headhunting, or have a job advertisement for popular positions, with often 500-1000 applicants applying for a single job.

To meet this challenge, recruiters within many firms and organizations use profiling to screen applicants for the most promising candidates. Typically it is used as an initial big cut, where some candidates later will be interviewed and evaluated in more detail. Today, this initial big cut of candidates is often made "automatically", by an algorithm, with the selection based on a set of values. This is a technological feat that simplifies the life of a recruiter tremendously, and really can be seen as a necessary advancement to meet the conditions of hundreds and thousands of applicants for single positions, in today's job market.

What companies' uses are more specifically is data mining: machine learning with

algorithms and statistical learning. This if often referred to as using "Big Data", or "People Analytics"; i.e. using large datasets to identify parameters that represents the best candidates for various positions. The motivation for companies is often the claim to become more *objective* in their assessment [1, 2]. They can also be claimed to want to become more *certain* in their decision making for hiring. But what consequences does this reliance on technology have on the process of selecting and discarding applicants?

Let us start by taking a step back; if there are no other means at the recruiters and managers disposal to make judgements, they will have to rely on their *intuition.* Field-expert's intuition can very well be very good and accurate for making estimations within their specific area of expertise, but they will nevertheless be estimations based on subjective value assessment, and at times faulty. There is also the problem of *discrimination*, even without conscious intent. Today this is controlled by equal opportunities law; it is illegal to (explicitly) discriminate based on categories such as skin color, ethnicity, gender and age (and to some extent avoid also implicit discrimination). Yet, even with such countermeasures in place, discrimination seems to remain. As Pager and Shepherd exemplifies: "today discrimination is less readily identifiable, posing problems for social scientific conceptualization and measurement" [3]. This is often explained as the unconscious phenomenon of implicit bias.

1.1 Implicit Bias and Predictive Machines

More generally, implicit bias is often referred to as unconscious associations. This is what could lead to so-called *stereotype threats*; to attribute certain abilities and predicted behavior to generalized categories of people. For example, several studies have examined how different individuals judge a written application, as a resume or CV [4, 5]. The only independent variable in the studies has been the name of the applicant; either native or foreign sounding, alternatively a name associated to a minority group. In USA as well as in Sweden, an applicant with a native name is much more likely to be called to an interview, than one with a foreign name, even if their resumes were identical, and their merits therefore should have been identical.

There is a similar effect for hiring women to management positions. An illuminating example of the effect of how implicit bias could work is the following:

> "There is good news and bad news about actual gender-related managerial differences. The good news is that some do exist. The bad news is that they are overused as the basis for sexual stereotyping." [6]

This might be a controversial finding, that there seemed to be a difference between female and male performance related to management positions. However, the authors also note that the difference is overused in recruitment. This is very much what can be seen happening when over-categorizing. While there might be a difference in performance, the difference is neither black and white, nor clear-cut. Yet it continues to affect diversity, and remains as an obstacle for diversity and equality.

I will exemplify this further with a more concrete example (see subsection 3.5 about predictive brains), but let us first get back to the machines and algorithms that try to predict the likely performance of applicants for a job position. Barocas and Selbst [2] argue that an analysis and an algorithm are never better than the dataset on which they are based; if the dataset is biased, so will the result be. Data sets are, just as in empirical science, historical data, which may themselves be laden with implicit biases of a discriminatory nature. Thus, even if there is no explicit rule or class inferences made in the algorithm to look for sensitive categories, the result from the algorithm may still be (implicitly) biased.

One example of this is in plain view; our language and our use of lexical words. Within our semantic use of words, implicit association tests (IAT) show how we, for example, associate the word "dangerous" more to "black", rather than "white" (mainly in a context of US). This can feel troublesome, if you discover an implicit bias in yourself, and there are several IAT's on the web that anyone can use to

investigate their own implicit biases. When it comes to machine learning and algorithms, a standard use is keyword-associations, and pure statistical learning. What Caliskan-Islam et al. [7] show is that when running IAT's on algorithms such as these, they show similar kind of implicit biases as humans do. It makes sense, if you accept Barocas and Selbst [2] message that an analysis is never better than the data set, and if there are (historical) biases embedded, as there seems to be in language, the results of this kind of machine-learning algorithms will also be biased.

But, discrimination is punishable by law, you might object, and has been for quite a while in most western countries. However, as you may already have realized from the section above, the use of algorithms is technically quite complicated, and their effects are not at all transparent. Therefore, there are currently holes in the available laws, allowing some types of biased machine-based profiling, for example of applicants and employees. I will discuss if better laws could solve this, such as those that are under way in EU (see subsection 4.2), and I will also discuss solutions involving "better" algorithms and machine learning techniques (see subsection 4.1).

Solution with human thinking. My main claim in this paper, however, is that it is necessary to look deeper on the human side of the equation. Behind all the use of algorithms and computers, there is at least at some stage, a human considering a selection of applicants, or making a decision. Another approach could be to try to stimulate better thinking, with the use of assistive technological tools, and I will make an initial exploration if lessons from methodology of science can be used. As mentioned above, when using machines and data, there is often a perception that you gain certainty and objectivity; cold hard facts, if you like, that are mined from the ground like minerals [8]. Science, however, would hold that what you need is *uncertainty* and *doubt*, and to always be open minded to question, critique, and try to falsify what you think is true. Empirical conclusions in science are, in other words, never facts set in stone, but always changing and open to new interpretations and new explanations.

1.2 Methodology and disposition.

There is an inherent problem in this sphere of investigation: companies using Big Data and data profiling. Namely that currently there are insufficient laws to force them to disclose their processes, or algorithms. As Frank Pasquele names it, we currently very much live in "The Black Box Society" [9], where we do not know what algorithms and technologies are affecting us. In lack of empirical data concerning the effects on society, what are left are either journalistic endeavors trying to peek into those black boxes, or theoretical works, looking at the mechanism of the algorithms that we at least to some extent know to be at work in the background.

The latter is my approach in this paper, and in section 2 I start by looking into the mechanisms within machine learning and data profiling. This is continued with an analysis of potential problem, in section 3, and a continued discussion follows in section 4, in which I also propose different solutions and remedies to tackle the problems at hand. The methodological limitation of limited empirical data in this sphere is addressed more in the ending sections.

2 Organizations using Big Data

The term *Big Data* is fairly new, but the phenomenon behind it is not necessarily as new. Broadly speaking, it concerns the acquisition of useable, meaningful data, out of larger data set, with the emphasis of the scale of the data set, or multiple data sets. This can also be called *data mining*, and they remain closely related [8]. Another term that is ubiquitous within the domain is *machine learning*, which focuses on using algorithms that "learn" to identify patterns for classification, from training data. Data Profiling is a more specific term within data management, and is defined as: "a specific kind of data analysis used to discover and characterize important features of the data sets. Profiling provides a picture of the data structure, content,

rules and relationships, by applying statistical methodologies to return a set of standard characteristics about data." [10].

In other words, what you get and can use from data mining and profiling is much like data from empirical, statistical research. What the different fields also have in common is that they mainly look for patterns in the data/training set; in much the same manner as a speculative scientist could do to generate hypotheses from empirical data.

The usage of big data and data mining is widespread. According to a survey in North America in 2014, 73% of companies stated they had invested, or planned to, invest in Big Data analytic tools [11]. Larger employers today are, almost, forced into using some kind of Internet-based management systems for handling job applications, either themselves or through an external agency. In such systems, they typically use Applicant Tracking Systems (ATS), to electronically manage information [12]. From this, it is a small to also rank applicants, and typically present only the select few applicants that are deemed to be the best fit, for recruiters and managers. The rest are, in other words, disregarded and never considered by anyone but the machine and the ranking system. Below I give a glimpse into the kind of parameters that can be used in ranking systems like these, with an example of Xerox looking for customer service personnel.

2.3 Example of Xerox using Data Profiling

Xerox is a global business services company with over 130,000 employees within various technological fields, as well as organizational development. An important element in their business model is customer service support. To manage their human-resources within this field they have implemented big data analyses. One explicit performance value that they look for in a candidate is *longevity* within the company, as in how long a person stays within the employer [13]. The motivation is often that longevity of workers builds organization-specific knowledge over time, and it is important that this stays within the company. As an analytic tool they have been using Evolv, currently part of a larger platform called Cornerstone OnDemand [14].

Xerox state that they put applicants through a series of tests to predict how they would perform within customer service [13]. This is later followed up in regards to performance measures, such as longevity. Evolv claim to have "500 million points of employment data on over 3 million employees" (see Christl and Spiekerman [15] for a more detailed overview of software used in hiring today). Part of the tests are surveys regarding attitudes towards work, such as how long they are willing to travel/commute to work, and how much overtime per week they would be willing to work. Additional information is also gathered from CV's.

Four measures were correlated with longevity: (1) Willingness to work 1-3 hours overtime per week, 15 times more likely, (2) Applicants with bachelor's degrees stay 5% longer, and those with technical diplomas stay 26% longer, than those with high school diplomas, (3) Those that have had a customer service job where they have had to use empathy, rather than just taking orders, and, finally, (4) living closer to job, or having reliable transportation [13].

It is also claimed that their ideal customer worker is someone that (1) scores well at typing tests, to be better at taking in background documentation on clients, (2) is creative, and (3) uses social media (but not too much). It is unclear how this data is correlated to performance [13].

The result here is a profile, with a set of variables and parameters (a pattern), which human-resource management use as a predictor of performance and how well an applicant fits for the position considered. In other words, someone that lives further away from the working place, only has a high school diploma, and does not use social media very much or at all; ends up with a low predictor and is likely to be cut at an early stage in the hiring process.

Some also propose using data mining profiles together with expert systems, to help make the selection of applicant [16]. This could be done with more or less automation, either by giving suggestions to a recruiter (a suggestion the recruiter is probably likely to follow), or by the system itself making initial selections. As

mentioned above with ATS's, this process of profiling was previously more manual, and is today highly automated.

No example has been found of recruiters using content data extracted from social media to include in a job profile predictor. However, a study in Sweden showed that at least half of the interviewed recruiters in big and middle-sized organizations did scan applicant social media profiles themselves, at some point before hiring [17]. In the survey they stated they looked for risk behaviors, where the applicant shared the organization's values, and things that might damage a company's reputation. It would not be a difficult step to include data from publicly, obtainable, social media profiles of applicants, and make a keywords-association analysis to (try to) capture potential risk behavior.

This gets us into an even more intricate sphere of privacy and the use of personal data, and it will expand the ways in which a person can be unfairly disregarded and discriminated. However, it may be enough to look at the data that an applicant has to include in a resume, as I will argue for in this paper.

On the face of it, when it comes to unfairness, social exclusion and discrimination, it is important to look for diversity, and if you like, uniqueness. This term is also used in data profiling, which I will consider next.

2.2 Uniqueness in Data Profiling

An important concept with data profiling is *uniqueness*. Technically understood, one value for a parameter, or a set of values for a set of parameters, has maximum uniqueness when there are no other sets of values like it. Non-uniqueness is achieved when two or more sets of values are all the same, which is the strongest kind of pattern and correlation.

Translated to a dataset of applicants and their level of academic degree, it would mean that maximum uniqueness is achieved when every candidate has a different level. On the other hand, if, for example, a test of programming skills is included in the dataset, maximum uniqueness would be when every individual has a different test score. This might become a bit more complicated when you have numbers with decimals, and in general higher uniqueness when there are a greater number of possible values. The principle, however, remains the same: that minimal uniqueness will be acquired when (groups of) individuals are (nearly) identical.

Taking the same dataset and looking at a relation between variables of programming skills, as a performance measure, and academic degree, *minimal uniqueness* is achieved if all individuals with high programming skills also are at a *certain level of degree*. This would make these easy to identify; just by looking at their academic degree it would be possible to tell something about their programming skills. On the other hand, if programming skills and academic degree are *not well correlated* but varies greatly, with good and bad programmer within all degrees, then *uniqueness is high*.

What does this mean for a recruiter who is looking at applicants and wants to find the best fit, given the motivation that you want to be as certain as possible? On one hand, you would like to be able to make the clearest distinction between, for example, a high and low performer. That you would get by having low uniqueness: if all low performers have a high school diploma, and all high performers have a bachelor's degree, the inference is easy to make, solely by looking at academic achievement. On the other hand, having high uniqueness will give you a much more narrow set of candidates that may be the *best fit*, and be able to disregard a larger portion of applicants. The latter likely depends on having a profile with a specialized pattern of skills and parameters.

3 Analysis of Potential Problems

Using profiles to cut a large portion of applicants might be seen as an efficient way of siphoning out the candidates less likely to perform well, and consequently a mere "fair and square" competition between applicants. Overall, employers and companies

should be able to perform better, with more people in the right positions. Companies like Xerox also report results going upwards after using data profiling for hiring [13].

I will start by analyzing some potential problems in the use of data profiling as described above. Initially, there might be a problem of effectiveness, and long-term performance, and I will use the conceptualization of local optimums. Secondly, and thirdly, I will get to the problems that are of more concern in this paper; regarding discrimination and unfairness. Further on in this section I will look at some initial technological remedies to the problems, with re-profiling, which seem to deepen the problems at hand.

3.1 Local Optimums and Stagnation

An ongoing research question within product development concerns innovation and success in different economic contexts. To shed light on this matter, some have applied an evolutionary perspective, and something called a fitness landscape [18]. This is related to the theory of evolution, in the sense that you make an analogy to which "ideas" will survive and progress, and will have the best fitness for a specific environment. Within this concept you can also make a difference between local and global maximums, and this related to what machine learning call optimums as I will henceforth use. To explain it further; an organism, with certain attributes at its disposal (a certain genotype), adapting to a certain environment, is theorized to sooner or later end up on a local peak (performance) optimization, for that specific environment and context.

Related to humans and organizations, this could mean that given a certain type of workforce you will end up at a certain level of performance. Within your own workforce you should with a large enough dataset also be able to observe different peaks. This could probably be done in larger organizations with (big) data analyses, or large collections of data points as in the program Evolv and Cornerstone (see subsection 2.1). You would then typically want to identify where the peak performances are, and spread *the type of abilities* that those people have to other positions in the same occupation. To put it differently; you would like to have the profile of peak performers in all relevant positions, given the current conditions.

However, when changing to a different environment, the same set of optimized abilities may not be as effective or competitive, compared to other sets of abilities. The same thing happens in the natural environment with highly specialized organism. If you change their environment, they might not be able to adapt, and dies out. The data that is used to make job profiles, like all data in empirical research; it is historical data. They do not really say anything about the future, and they only say what worked well in that historical context/environment.

This could be problematic, given that the global market, influenced by people and societies all around the globe, can involve quickly changing conditions for success. This is a problem for any use of profiling, of course, not just data mined. I will discuss a potential solution with re-profiling below (section 3.4), and get back to this problem in section 4.

3.2 Problem with Discrimination

The most ethically relevant problem is probably that of discrimination in the selection process. As mentioned in the introduction, discrimination is controlled by legislation in most countries, and is usually defined as: "the treatment of a person or particular group of people differently, in a way that is worse than the way people are usually treated" [19]. Expressed in a different way, it is about judging people based on things like social group categories, rather than individual merit. Most commonly this is an issue about features such as: (1) gender, (2) ethnical background, and (3) sexual orientation.

It is important to note that discrimination can be direct and based on discriminatory categories, such as those mentioned. It can also be indirect; a

selection based on non-discriminatory categories, which, however, are strongly correlated to the discriminatory ones.

There are, for example, reports of companies scanning employee's health records, to predict those likely to get on sick-leave, or those that might be pregnant [20]. Why this is possible, and not an offence, currently, is highlighted in discussing solution with law and regulation (see subsection 4.2).

3.3 Unfairness and Loss of Opportunities

A more general problem and much related to discrimination, is simply *unfairness* in the selection process, and a *loss of opportunity* for those who deviate from (outliers) the resulting profile. That is, this is the individuals that differ from the norm of the (predicted) most likely high performers, but who in reality still is just as likely to perform at the same level. This is likely to occur, given the premise that measures for the applicants have statistical normal distributions. To use the example of Xerox (see subsection 2.1), it is unlikely that *all* applicants with high school diplomas would perform worse than *all* the applicants with a bachelor's degree.

This is to some extent related to uniqueness in data profiling. Given that if the more informative measures related to performance will have lower uniqueness, they will also be less forgiving towards outliers. Those outliers could be seen as losing an opportunity, and this loss could be an ethical problem of *social exclusion*.

It is important to note, though, that we would probably not perceive it as unfair if other candidates in fact had stronger merits, but it would be perceived as unfair if you are excluded based on something not directly associated as a merit for the position. Xerox was, for example, looking for people with mid-range social media use, for a position in customer service. Using social media is not, at least, directly related to the task that they are expected to do, but apparently it might be indirectly related to a performance measure. It could possibly exclude applicants solely based on not using social media, or not using it enough. Given that people are adapting to the demands of the job markets they want to get into, it might mean that people will have to conform their private life, as well.

3.4 Reprofiling for Quality of Data

One approach to remedy some of the problems just mentioned might be to simply seek better quality of the data. It could be claimed that the problems of local optimums could be avoided if the profiles were continuously updated to reflect what kind of workers are needed at a specific moment in time. Even more so, if the quality of the (meaningful) data extracted from the big data set is good enough, and detailed enough, it should be possible to pick out those individual outliers who nevertheless are performing at a high level.

This does not seem to be an unreasonable claim, but it might be problematic in practice. As I will show further down, the process of applying a profile might reduce all the outliers. I will also relate to this process to intuition and implicit bias in the next subsection.

Data quality is first and foremost an issue when it comes to, so to speak, living data sets that change over time (such as what correlates with performance within a certain organization on renewed global markets). To remedy this, most data sets can benefit from what is called *reprofiling;* a new analysis of the dataset to see if statistical patterns and relations have changed [10].

However, a risk, as far as I can analyze it, is that rather than including outliers it might exclude them further. Let us say that we mine out a profile P1, with features such as those that Xerox are looking for in customer service personal. You then choose people based on profile P1 to be added to the workforce for the specific occupation, and then re-profile on this new population with the new selection of individuals added, yielding a profile P2.

If nothing significant has changed in regards to what measures are correlated to high performance between P1 and P2, the added employees will have reinforced the

pattern, which statistically will become “stronger”, as in, having less deviations from means, and lower p-values. This will most likely mean that P2 would be even less forgiving to outliers, unless your data analysis is detailed enough. Big organizations like Xerox with abundant resources at their hand might be able to make the profiles detailed enough, but would probably be difficult for others.

Revisiting Local Optimums. So, what does this then mean for the problem of ending up on low local peaks of performance in changing market-environment? If profiles P1, or P2, have had any impact on the workforce, you would after some time have few outliers to rely on. With fewer outliers it will be more difficult to find the pattern for peaks of performance in a new environment. The workers who would be better suited for this new environment might simply not be left in the company.

On the face of it, then, reprofiling might not be enough to solve the problem of local optimums. I would like to note three premises for this effect to be relevant: (1) profiling and reprofiling is made on a closed group, as in not including a larger part of a specific market environment, (2) that using profile P1 actually makes the pattern P2 have lower p-values, and stronger correlations with performance measures, and (3) that the analysis is not focusing on specific, high performing outliers, but rather use the profile for generalized big cuts of applicants, like Xerox seems to use it (see subsection 2.1). Otherwise the theorized effect would not be there.

3.5 Human Intuition and our Predictive Mind

As already has been touched upon in the introduction, the problem with profiling and reprofiling might not be too different from what we initially wanted to get away from; the human, intuitive way of reasoning and categorizing. The data could be biased, implicitly towards certain categorizations. What I want to propose here is that profiling and trying to predict future high performance might be even more directly related to human cognition.

An explanatory theory within neuroscience that has gained momentum lately propose that the brain essentially should be seen as a hypothesis-testing mechanism, that performs a kind of statistical analysis within the neurons [21]. From this perspective, the main goal is for the brain to be able to predict what will happen, before it happens. Within the connections between the neurons there is both a forward progressive movement; receiving input from nerve-endings, for example our eyes, to be processed in succession in the neural cortex. But there is also a backward progression, where signals are constantly being sent backwards, from higher-cognitive areas of the brain, towards lower. This has previously been explained as a kind of feedback-loop, but lately the backward progression has been given more emphasis and use to explained human cognition the other way around.

Rather than starting blank and use input to create a model of the surroundings, the human brain starts with a working model of the world that is projected backwards within our connective network; and ultimately onto the world around us. The signal input functions as an error-correction function to adjust our working model of the world, where it is needed in order to explain what is happening in a specific moment. Accordingly, no corrections of our model will be made, as long the model can explain what is going on around us.

Studies also show that the kinds of categorizations we do to learn about the world are very much like statistical analyses. We more or less make a *mirrored casual structure of the world* in our brain, based on what we have experienced [21].

All this can be related to something called *confirmation bias*, which means that humans have a tendency to direct attention, unconsciously, towards aspects which confirm our prior beliefs. In relation to the theory of our predictive mind, this would mean that in a backward loop we project a working model of the surroundings, and the attention is drawn towards that which confirms this model to be true. In a sense then, our prior beliefs and categorizations made from experience, determines what kind of information we will look for when facing a new situation; that which confirm what we already know, or believe that we know. Only when there are too much obvious errors, we start to look for a different explanation, and correct the working

model. It is then also important to note that often there is insufficient feedback to correct our model when it comes to valuation of people, or groups of people, that we encounter. In the absence of sufficient errors, the working model will be reinforced; as functional to predict the outcome.

In some sense this can be related to profiling and reprofiling. After making a categorization, like for a profile, it will direct attention only to the feature you know they are related to. Even when attempting to reassess that knowledge, as in if you re-profile a population that has been modified by the previous profile, it is likely that you will end up looking at the same features. These features have then also become statistically stronger, as was hypothesized in the previous subsection and that will further reinforce your prior belief. A risk with human intuition is that prior beliefs dominates, controls your attention, and determines your future beliefs; the same risk can be found with reprofiling in data mining.

Revisiting implicit bias. What seems to happen with the described function of our predictive mind is essentially what is happening with implicit biases; namely an attempt for the brain to predict future outcomes based on previous experience and working models. As already mentioned, recruiters seem to try to predict performance of applicant, solely by looking at their names (see subsection 1.1, [4]). Implicitly, we seem to have working models with rough estimations, based on categorizations of people.

The delicate problem of this can perhaps be emphasized more clearly with an example of gender difference in math-skill. SAT-results in USA from 2013 show that male students performed better; an average of 531 (SD = 121) for male students, versus 499 (SD = 114) for female students ($p<.001$) [22]. However, given the standard deviations, and normal distribution around the mean, 38% of female participants outscored 50% of the male participants (this is illustrated in Fig. 1).

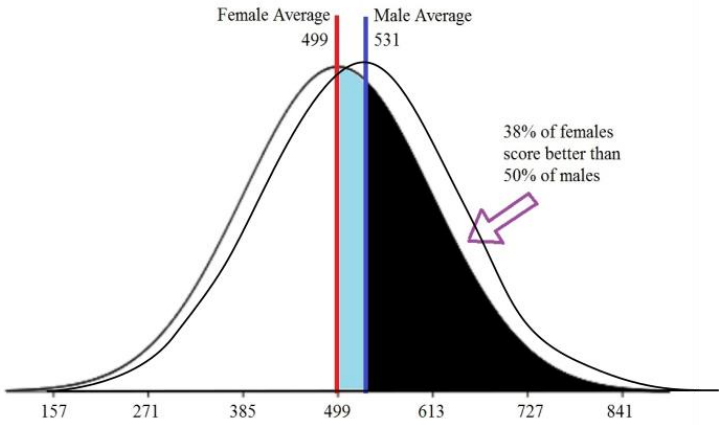

Fig 1. Division of SAT math-test scores of 2013 in US [22]. Even if there is a significant difference in average score between genders ($p<.001$), 38% female student will still outperform the average male participant, statistically. This is represented by the black part of the graph.

Faced with one male and one female in front of us, and the task of selecting one we deem most likely be best in math; then the difference in score is not very informative. There is almost 40% chances to be wrong if choosing the one associated to the group with the highest mean. That is not very good odds, considering that 50% is pure chance. Despite this, most people will feel confidence in their intuition in cases like this, and gender based stereotyping and over-categorization will therefore remain. If this is anything like implicit biases it will also be difficult to get direct feedback on this implicit bias to adjust our belief, to at least not be as certain in our categorization.

This is ultimately a limitation of statistical significance measures, that even if you have as strong of a significance as you basically can get ($p<.001$), you end up with a “meaningless” knowledge. This has, of course, also been recognized by others , and a

way to counter that limitation is to look for *statistical power,* or *importance;* more specifically to include a measure of effect size [23–25].

Analyzing effect size, the gender difference shown in the SAT math-test can be concluded to represent only about 3% of the difference attributed to gender, and the other 97% attributed to other factors (Choen's d = .37, effect size r-squared = .03) [22]. This would be another way, a statistical way, to say something about how informative a significant difference is, as in the situation of choosing between two applicants of different gender based on average SAT-math score.

Revisiting the customer service profile. Making an analogy between the example of SAT-math score above, and the profile measures from Xerox, it may be significantly more likely that applicants that live closer to work will on average perform better, and stay longer at the company. However, some applicants living further away might be just as likely to perform well at the job. Nonetheless, if all other parameters are the same, the one who live further away will likely not be presented to the managers that hire the new customer service personnel.

4 Possible Solutions

What I have argued for so far are potential risks for using data profiling when hiring. I have also tried to show that these are not necessarily any new risks, but can be seen in any use of profiles. It even seems to relate to the way our human intuition categorizes from experience; we try to predict future outcomes just like algorithms.

There may be different ways to try to attack the problems highlighted in the previous section; i.e. local optimums and stagnation, discrimination and unfairness. I will recapitulate some previously discussed approaches, mainly; (1) create better and more accurate algorithms and technology, and (2) introduce more extensive laws and regulations. I will add to the discussion a third approach: (3) promote better thinking for the individuals who, at some stage or another, still will be involved in the decision and selection of applicants. They are not intended to be mutually exclusive measures; on the contrary, all could contribute and be equally as important for a sustainable solution.

4.1 Solving it with technology and better algorithms

An initial concept to consider is that of overfitting in machine learning. It is a term related to "fitting a model" to a training set of data, as in trying to explain the dataset, its relations, and its values. In overfitting, the statistical model has included irrelevant features, random error, or just noise, rather than only valid relationships and correlations. Consider the SAT scores discussed above; adding an analysis of effect size reduced the importance of the significant result of gender difference. You could do the same analysis of statistical power, such as looking at effect size, on a model in machine learning. What you can do after that is something called "pruning", which means, analogically, to cut off the branches of the decision tree that lack predictive power (like gender difference in the above case), which will improve the overall predictive power of the model (or pattern).

There is reason to be a bit skeptical whether this is a solution that will be implemented in practice. In psychology research, a standing advice for 25 years has been to include analyses such as effect size. Sedlemeir and Gigerenzer [24] could conclude that only 2 out of 64 experiments even mentioned power and effect size, and instead solely relied on statistical significance. Another meta-analysis concludes that statistical power has not improved significantly in the 60 years since the concept was introduced [25]. Thus, it is likely even more difficult to get employers and companies in the job market to use measures of statistical power instead of relying on (simple) significantly correlated relationships.

Including Diversity. Another way of solving the problem of discrimination could perhaps be to directly include diversity. Potentially this could also solve problems of

stagnating at local maximums. If conditions in the environment change, a diversity of skills in the workforce would make a company more likely to be able to adapt. But is diversity enough to remain innovative and adaptive? Some research seems to suggest just that, where one study finds a correlation between occupational diversity and likelihood of innovative research [26], and another study finds that openness to cultural diversity was correlated with the likelihood for a company's continued economic performance, which was interpreted as being able to renew itself and to innovate [27].

Whether it would be a solution against discrimination might depend on in what categories, and how, you create diversity. However, you could at least try to account for actual differences, rather than overgeneralize. For example, let's say we have a specific task to find someone solely based on math skill. Related to the example of gender differences in SAT-math score in subsection 3.4, implementing diversity could be to, at least, have result distributions similar to the test scores. In other words, you would want to see about 38% female and 62% male candidates, instead of over-generalizing to only look for male applicants. Just any man will not be the most likely best mathematician.

The same thing, then, could be applied to the relation between academic degree and customer service workers within Xerox (see subsection 2.1). Without having specific numbers or distributions available, but based on the assumption that these will follow normal distributions, you would want to, at least, see a similar spread and diversity in the applicants that you accept. Those with bachelor's degree will be more in numbers, but not exclusively.

This could also be a possible way to open up for equal opportunities for people that are just as likely to perform well within a certain occupation, that otherwise would not have gotten the chance. In a very direct way the purpose can be said to be to not limit uniqueness too much in your data as well as with individuals, when including diversity.

If this is the case, that diversity is wanted, you could ask yourself why profiling at all. Why data profiling at all, if it is good to have a diverse workforce instead of one with certain, highlighted, high performing features? To some extent it might be a question of balancing between long-term and short-term goals. On one hand, the data probably doesn't lie, and if a company *adopts* the profile they get from current data, the company would probably perform well in the near future. On the other hand, it might become difficult to *adapt* to changing conditions, and will not perform as well long-term.

Including anonymity. Another technical solution to unfairness in selection processes could perhaps be to make applicants anonymous. This is an often proposed remedy for discrimination; to anonymize resumes [28]. Technically, it could for example be to use k-anonymity [29]; that a person is anonymized with k-other people. It could mean that applicants are anonymized with all the other (k-amount) applicants for the position; as in, removing all personally identifiable data, like names and address. It remains a problem that smart algorithms can correlate impersonal data to other databases, and still figure out who is who, but this remains a separate problem.

Anonymizing could then be a solution to unfairness, to not include categorizes of people that are laden with pre-conceptions in the evaluation and prediction of future performance. However, it seems less of a solution for some of the problems discussed so far in this paper. Like the use of academic degree, and other data that inevitably must the included in a resume and CV.

This can be discussed further; what should be considered as merits for a position? Basic qualifications should be met, of course, but should it always be in terms of an academic degree for example? Working with telephone marketing, or customer support, may have a performance measure related to academic degree, but is it really necessary? This is a bigger discussion that does not fit within the limited space for this paper. More importantly, discussed below is the use of laws that effectively could function as anonymization if the use of personal data is prohibited.

4.2 Solving it by Law

The skepticism mentioned above, regarding wether companies would really implement certain solutions, could perhaps simply be regulated by law. Currently, laws are not proficient to hinder (potentially) harmful profiling. Barocas and Selbst [2], for example, examines anti-discrimination laws in the US, and conclude there is limited support if there is no (conscious) intent involved. The "Equal Employment Opportunity Commission Uniform Guidelines", as well, does not seem to be able to hinder such analyses. Rather they seem to explicitly allow the necessity for predicting future outcomes of recruitment and employment; very much what profiling on the surface is doing [2].

Crawford [30] also notes that Big Data systems seem to have similar problems as many governmental administration systems have; like the lack of notification for affected individuals. The proposed solution is to, at least, notify an individual potentially affected to "predictive privacy harms" [30]. And if they do not agree with the use of their personal data, individuals should be able to retract it; also from the predictive algorithms that does not affect them personally.

In just a couple of years, 2018, EU will incorporate a new set of rules; the "General Data Protection Regulation" (GDPR). In it, there is a specific section (Article 22) for decision-making of individuals, including profiling. Data processing is defined as profiling when "it involves (a) automated processing of personal data; and (b) using that personal data to evaluate certain personal aspects relating to a natural person" [31]. In other words, it does not have to be yourself that the usage of your personal data affects. This would often be the case in the examples used in this paper, with personal data from previous applicant and employees that is applied on new applicants. Personal data is defined as any data that relates to a person's private, professional or public life; like a name, photo, posts on social networks, or your computer IP address.

This seems to be quite extensive, and my analysis will not go into detail of how effective they will be put into practice. I will merely note that there is an exception clause if you give consent to its use. You are still able to retract it afterwards, but on the face of it, there seems to be an opt-out implementation; as in, only those that actively opt to restrict their use of personal data, like in a consent form, will hinder the use. It is not likely a majority of people will opt-out of this kind of use, and there for there will probably still be much data to be data mined. The transparency of when you have been affected by an algorithm-based decision, is not straight forward; unless, perhaps, if you do not also adopt the notification principles as Crawford [30] proposed above.

4.3 Promoting better thinking

The last approach would be to more explicitly look at the human side of the decision process. Will regulations and technical opportunities be enough to guide human behavior, towards fair and non-discriminatory behavior? Laws and technology can create restrictions and constraints for what kind of behavior is likely to be performed, but if profiling in an unrestricted fashion remains profitable, it is likely many will find a way through the blockades, so to speak. It would demand further analysis if it would be enough to change recruiter's and manager's behavior, and at this point I will merely remain skeptical. This skepticism can neatly bring us to the main argument in this section.

What recruiters and managers often seek when hiring, is support in their decision; they want to be more *certain* of their choice [1, 2, 16]. Similarly the brain wants to be more certain of the future, predicting what comes next. As was touched upon in the previous section, the way our brain functions in this regard may also lead us a bit astray at times, leading to implicit biases and negatively value laden pre-conceptions. What might be needed, then, is a bit more *uncertainty*, and a motivation for this can be found in the methodology of how *good science* is performed.

A key conception in modern empirical science is to *not* try to confirm what you believe to be true, but to falsify it. This relates very much to our confirmation biases

and our tendency to look for that which confirms our prior beliefs and current working model in our brain (see subsection 3.5 for details). What the methodology of science does is, instead, to critique and to doubt, and try to disprove what we believe to be true. As Kothari puts it: "All progress is born of inquiry. Doubt is often better than overconfidence, for it leads to inquiry, and inquiry leads to invention" [32]. A similar message is delivered by Nobel Prize awarded physicist Richard Feynman:

> "We have found it of paramount importance that in order to progress we must recognize our ignorance and leave room for doubt. Scientific knowledge is a body of statements of varying degrees of certainty — some most unsure, some nearly sure, but none *absolutely* certain." [33].

So, rather than certainty, what you would like for scientifically more sound knowledge, is a promotion of *uncertainty*; or varied degrees of certainty and uncertainty, based on empirical data and arguments. What a scientific approach to data mining for applicant prediction could mean is that you as a recruiter should actively try to disprove your own hypothesis regarding who you believe is the best candidate for the position.

5 Conclusions

The main purpose of this paper is to present an initial, explorative investigation of ethical aspects within recruitment and hiring using Big Data and data profiling. It is found that the use of data mining in companies is often motivated as an attempt to increase objectivity in regards to performance predictions. As far as has been analyzed here, there seems to be some potential problems that companies risk running into. The first is to end up on low peaks of local maximums and therefore miss out on achieving long-term performance gains and resilience against changing conditions. The second is a concern for discrimination when using data mining algorithm. The third and final, and perhaps a sometimes forgotten focus, is a concern for loss of equal opportunities and unfairness for applicants in a selection process.

The problems discussed here are not new, in two ways. First of all, the problem of biases in data mining and data sets is well known, and even more specifically, implicit bias has been shown to exist in the use of data mining [2, 7]. Secondly, problems with discrimination and (human) biases are well known and studied, and as I propose in this paper, these phenomena seem to be related in more intricate ways than perhaps often considered. Once again relating to the prediction-error theory of our mind (see subsection 3.5), data mining and profiling is an attempt to predict the future, in a similar way as the brain is trying to predict the immediate surrounding and what will come next. An inherent limitation in data mining and the ability to predict the future is that it is based on historical data. Similarly, the brain's predictions are based on historical experiences, and the brain can also be understood to do a kind of statistical analysis, leading to estimations and generalizations. This makes the brain prone to faulty predictions and discriminatory categorizations of people, and similar effects can be seen in statistical methods of data mining.

To continue on that analogy, it could be seen as if (big) data profiling, purely by its statistical method, can have implicit biases. Alternatively, it can reinforce the idea that humans through the perception of the world are (imperfect) statistical machines.

In this paper I have briefly touched upon some previously discussed ways to remedy these limitations in data mining and machine learning; through technological means (see subsection 4.1), or by laws and regulations (see subsection 4.2). A third option proposed here is to look at the quality of the human thinker and decision maker, and more precisely what could be the remedy in the scenarios of discussed with recruiters and managers. If it is better knowledge in our predictions we want, the proposition is to look at scientific methodology (see subsection 4.3). In that case, rather than looking for objectivity and certainty in decisions and selections of applicants, we should look for uncertainty, and doubt, and try to falsify our initial judgements of applicants ability to perform well.

How this openness to uncertainty, and a varied degree of certainty, is not either so

clear. These are abilities that can relate to critical thinking skills, and how this is taught and improved is still an open problem without a clear answer. Even if we are able to define such a skill-set, and how to promote them using the predictive data-mining tools recruiters can use, it is an inherent difficulty why this would ever be implemented by companies doing the hiring. In the environment of big companies using Big Data, a common understanding often heard is that "all you need is correlations", with no need to go beyond that understanding (Roger Clarke in lecture at IFIP 2016, Aug 21st). As has been shown in this paper, using only significance and p-values as statistical mean, you might end up with limited knowledge, and at times knowledge that breeds implicit biases (see subsection 3.5).

To conclude; what might be needed to promote a more conscious use of predictive tools, then, is a mix of technology and laws that is grounded in an understanding of how human decision-making and prediction-making is made.

5.1 Further Investigations

As was mentioned in the methodological subsection (1.2), there are currently some obstacles in doing research in the field of recruitment using data mining algorithms. Presently, companies and agencies do not have to disclose what algorithms they use, and how it affects populations [9]. This paper remains mostly theoretical. To get more insight into the effects on individuals and society, researchers would need access to more information about what kind of algorithms are in use, and also data from populations (applicants) it is applied on.

Some of the things argued for in this paper could be simulated and tested. For example, the effect argued for with reprofiling; that statistical patterns would become more significant and with lower p-value, and consequently result in less deviation from means for individual features. Another effect to expect if lots of data points are added is that p-value will be low.

Finally, the discussion on what constitutes discrimination and unfairness would serve to be expanded. What counts as merits? When is it ok to use predictive measurements? Should they never be allowed? If they are statistical estimations and generalizations, they would inevitably lead to excluding some individuals. Should profiling be allowed if the statistical measure is good enough, and what constitutes good enough? These are some questions that do not seem to have clear and immediate answers.

Acknowledgements. I would like thank my supervisors Mikael Laaksoharju and Iordanis Kavathatzopoulos of Uppsala University, for extensive feedback and conceptual contributions.

References

1. Chien, C.-F., Chen, L.-F.: Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. Expert Syst. Appl. 34, 280–290 (2008).
2. Barocas, S., Selbst, A.D.: Big Data's Disparate Impact. Calif. Law Rev. 104, 671–732 (2014).
3. Pager, D., Shepherd, H.: The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. Annu. Rev. Sociol. 34, 181–209 (2008).
4. Bertrand, M., Mullainathan, S.: Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Am. Econ. Rev. 94, 991–1013 (2004).
5. Jost, J.T., Rudman, L.A., Blair, I.V., Carney, D.R., Dasgupta, N., Glaser, J., Hardin, C.D.: The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. Res. Organ. Behav. 29, 39–69 (2009).
6. Sexton, D.L., Bowman-Upton, N.: Female and Male Entrepreneurs: Psychological Characteristics and their Role in Gender. J. Bus. Ventur. 5, 29 (1990).
7. Caliskan-Islam, A., Bryson, J.J., Narayanan, A.: Semantics derived automatically from language corpora necessarily contain human biases. ArXiv Prepr. ArXiv160807187.

(2016).
8. Portmess, L., Tower, S.: Data barns, ambient intelligence and cloud computing: the tacit epistemology and linguistic representation of Big Data. Ethics Inf. Technol. 17, 1–9 (2015).
9. Pasquale, F.: The black box society: the secret algorithms that control money and information. Harvard University Press, Cambridge (2015).
10. Sebastian-Coleman, L.: Measuring data quality for ongoing improvement: a data quality assessment framework. Elsevier Science, Burlington (2013).
11. Gartner Survey Reveals That 73 Percent of Organizations Have Invested or Plan to Invest in Big Data in the Next Two Years, http://www.gartner.com/newsroom/id/2848718.
12. Rosenblat, A., Kneese, T., others: Networked Employment Discrimination. Open Soc. Found. Future Work Comm. Res. Pap. (2014).
13. Morse, T.: Big data can take the guesswork out of the hiring process, http://qz.com/146000/big-data-can-take-the-guesswork-out-of-the-hiring-process-2/, (2016).
14. Evolv | Cornerstone OnDemand, https://www.cornerstoneondemand.com/evolv.
15. Christl, W., Spiekerman, S.: Networks of Control: A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy. Facultas Verlags- und Buchhandels AG, Wien, Austrai (2016).
16. Mehrabad, S.M., Brojeny, F.M.: The development of an expert system for effective selection and appointment of the jobs applicants in human resource management. Comput. Ind. Eng. 53, 306–312 (2007).
17. Backman, C., Hedenus, A.: Will your Facebook profile get you hired? Employers use of information seeking online during the recruitment process. Presented at the The 6th Biannual Surveillance and Society Conference , Barcelona, Spain April 24 (2014).
18. Kwasnicki, W., Kwasnicka, H.: Market, innovation, competition: An evolutionary model of industrial dynamics. J. Econ. Behav. Organ. 19, 343–368 (1992).
19. discrimination Meaning in the Cambridge English Dictionary, http://dictionary.cambridge.org/dictionary/english/discrimination.
20. Wilkinson, J.: Companies secretly tracking employees' health and lives in "big data," http://www.dailymail.co.uk/news/article-3456691/How-companies-secretly-tracking-employees-health-private-lives-big-data-save-money-lead-big-problems.html.
21. Hohwy, J.: The Predictive Mind. OUP Oxford (2013).
22. Cummins, D.: Why the Gender Difference on SAT Math Doesn't Matter, https://www.psychologytoday.com/blog/good-thinking/201403/why-the-gender-difference-sat-math-doesnt-matter, (2014).
23. Kelley, K., Preacher, K.J.: On effect size. Psychol. Methods. 17, 137–152 (2012).
24. Sedlmeier, P., Gigerenzer, G.: Do studies of statistical power have an effect on the power of studies? Psychol. Bull. 105, 309 (1989).
25. Smaldino, P.E., McElreath, R.: The natural selection of bad science. R. Soc. Open Sci. 3, 160384 (2016).
26. Söllner, R.: Human capital diversity and product innovation: a micro-level analysis. Jena Econ. Res. Pap. 27, 1–33 (2010).
27. Østergaard, C.R., Timmermans, B., Kristinsson, K.: Does a different view create something new? The effect of employee diversity on innovation. Res. Policy. 40, 500–509 (2011).
28. Hutchison, K., Jenkins, F.: Women in Philosophy: What Needs to Change? Oxford University Press (2013).
29. LeFevre, K., DeWitt, D.J., Ramakrishnan, R.: Incognito: Efficient full-domain k-anonymity. In: Proceedings of the 2005 ACM SIGMOD international conference on Management of data. pp. 49–60. ACM (2005).
30. Crawford, K., Schultz, J.: Big data and due process: Toward a framework to redress predictive privacy harms. BCL Rev. 55, 93 (2014).
31. Heimas, R.: Top 10 operational impacts of the GDPR: Part 5 - Profiling, https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-5-profiling/, (2016).
32. Kothari, C.R.: Research methodology: methods & techniques. New Age International (P) Ltd., New Delhi (2004).
33. Feynman, R.P., Robbins, J.: The pleasure of finding things out: the best short works of Richard P. Feynman. Perseus Books, Cambridge, Mass (1999).