



LUNDS
UNIVERSITET

Lund University

Department of Philosophy

Bachelor thesis, FPRK01: 15 HP

What is Wrong with Implicit Bias?

An Investigation of the Moral Problems with Implicit Bias

Author: Jakob Olofsson

Supervisor: Dan Egonsson and Cathrine Felix

Time of examination: Fall 2014

Table of Content

Introduction.....	3
1 Implicit Bias	5
1.1 Introduction to Implicit Bias	5
1.1.1 Origins.....	5
1.1.2 The Concept.....	6
1.2 The Tests.....	7
1.2.1 Implicit Association Test (IAT)	8
1.2.2 The CV-error	9
1.2.3 Prestige Bias	10
1.2.4 Shooter Bias.....	11
1.2.5 Conclusions so Far	11
1.3 Preempting Criticism	12
2 What is Wrong with Implicit Bias?	15
2.1 What is Discrimination?	15
2.2 What Makes Discrimination Wrong?	18
2.2.1 The Dignity-based View	19
2.2.2 The Harm-based View	22
2.2.3 Implicit Bias and Discrimination	24
3 Summary.....	28
4 References.....	30

Introduction

‘Men hire men’, ‘we must fight structural racism’, ‘we must fight the patriarchal structures that dominate our society’. Sentences of this kind have been chanted as truisms after observing the unequal division between men and women in positions of power, the amount of non-Caucasians sentenced as criminals in relation to Caucasians – the list goes on. But the claims received vigorous critiques; ‘I am no sexist, I hire only by merit’ or ‘I am no racist, it just happens to be more criminals amongst non-Caucasians’. People who hire or practice law etc. believe themselves to be good people doing what is right. Until the 1980s, the structuralist theories were merely theories which, in Popperian terms, were unfalsifiable (even though they had high explanatory value). Since the 80s, it seems as our natural sciences have caught up with our fruitful, yet speculative theories and has introduced the concept of Implicit Cognition. Implicit Cognition could be understood as a process which is unavailable to us through introspection but infects our judgments and affects our actions. When we judge people, our judgment might in part be based on Implicit Cognition, without us having any knowledge about it. You might act in a sexist, racist or homophobic way without you even realizing it. This notion surely takes the edge of the critique based on introspection and good will. In the philosophical discourse, Implicit Cognition regarding social groups is named Implicit Bias, and Implicit Bias is the topic of this paper.

At first, the Implicit Bias might seem as a fulltime harmful process, but after further consideration it also seems to fulfill a non-problematic function. The Implicit Bias seems to be a process which helps us classify the world. Without such a process classifying would be very troublesome. In this paper I will investigate the ethical implications of Implicit Bias, i.e. I will attempt to answer the question: when is the process of Implicit Bias morally questionable and when is it not? I will in chapter 1 define the notion of Implicit Bias by discussing some of the contemporary theories by philosophers and psychologists. After reaching my definition of Implicit Bias I will argue that, when Implicit Bias is something bad, wrongful discrimination is its source of badness. To do this I will first in chapter 2 discuss what the nature of discrimination is and then further what is bad with wrongful discrimination. Due to the lack of space and time I will confine myself to discussing two views regarding the badness of discrimination, but I will argue that both views are possible

ways of describing the badness of Implicit Bias, when it is bad. To do this successfully I will develop the common view of what wrongful discrimination contains.

1 Implicit Bias

1.1 Introduction to Implicit Bias

In this section I will describe in short the origins of the concept implicit bias. I will also attempt to form a general definition of the concept which, in later sections, will develop to the final definition which will be used to investigate the ethical implications of implicit bias.

1.1.1 Origins

The philosophical discussions on Implicit Bias (IB) flow from the psychological findings regarding implicit cognition. In a range of tests, psychologists identified processes which were not the ordinary type of deliberative and conscious processes on which we, supposedly, base our behavior. Early discussions on implicit cognition and stereotyping can be found in Greenwald and Banaji's article "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes". In their article, they conclude that:

Social behavior is ordinarily treated as being under conscious (if not always thoughtful) control. However, considerable evidence now supports the view that social behavior often operates in an implicit or unconscious fashion. (Greenwald, Banaji, 1995, p.4)

To exemplify, consider Bob who edits a renowned philosophy journal. Bob explicitly thinks of himself as a professional, one who solely considers the work and the quality thereof without being swayed by anything but the rigorousness of the article considered. Despite his explicit professionalism, which Bob takes much pride in, he might implicitly associate Arabs with being dogmatic and fundamentalists. This association might influence how Bob judges articles written by Arabs and might cause him to refute articles written by Arabs instead of sending them to reviewers.

The early debate on how we are supposed to understand the term "implicit" can be divided into two camps: one identifies implicitness with automaticity and the other identifies implicitness with unconsciousness. The first stems from the distinction between "controlled" and "automatic" information processing. The controlled processing was suggested to be voluntary, attention demanding, and of limited capacity. By contrast, the automatic processes were hard to control, non-attention demanding and of nearby unlimited capacity. One can

only process small amounts of information in the controlled manner due to its attention-demanding nature. The automatic information processing can process much larger amounts of information since we are not required to actively deliberate on the matter. In the mid-90s, the psychologist Russell H. Fazio and his colleagues demonstrated that attitudes could be "activated" by either controlled or automatic processes. In this sense "activated" is to be understood as "caused by". Cases of the latter kind are often found when subjecting oneself to one of the more well known tests regarding implicit cognition, namely the implicit association test (IAT), which we will discuss more in section **1.2.1** (Brownstein, forthcoming)

The second way of understanding implicitness is by identifying it with unconsciousness. This idea is based on the fact that we have a hard time finding our biases through introspection. Consider Bob again, he believes that he is not affected by people's ethnicity and believes that he is acting accordingly, i.e. he is not at all aware of the ways his IB affects his judgments. Yet he is in fact biased and the bias might even influence his behavior. In section **1.2**, we will observe that there is often a discrepancy between what the subject believes to be his or her bias and what is shown by the test. This is thought to support the claim that implicitness is to be identified with unconsciousness (Brownstein, forthcoming).

1.1.2 The Concept

Initially, it is of great importance that we do not confuse the implicit with the explicit bias. When I say things like, "women love shopping", "American football players are stupid" or "boys like cars", I am expressing explicit biases. When Bob wants to hire a new associate, still committed to the idea that merit is the only thing that counts and tends to favor Swedish candidates over Arab due to his implicit association, he is hosting an implicit bias. To explain this difference, I will use the two terms discussed above, automaticity and unconsciousness because of the central position they claim in the discourse (granted that others might prefer different terminology). The implicit bias is a bias (we will return to the bias-term later) which is unconscious and automatic whereas the explicit bias is a bias which is conscious and controlled¹.

¹ An explicit bias is definitely conscious, it is more questionable if it is controlled in the sense above. I would suggest that even if it turns out that an explicit bias is uncontrollable it has the "feel" of being controlled. A person honestly uttering the words "boys like cars" is doing so believing that it is true and that their belief is somewhat justified. suggest that even if it turns out that an explicit bias is uncontrollable it has the "feel" of being controlled. A person honestly uttering the words "boys like cars" is doing so believing that it is true and that their belief is somewhat justified.

In contemporary discussion, it is up for debate to which degree implicit biases are unconscious (see Fazio & Olson, 2003 for further discussion). The yet to be published Stanford encyclopedia entrance on the matter defines implicitness as "relatively unconscious and automatic"(Brownstein, forthcoming). The author adds "relatively" to acknowledge the fact that it is an open debate. Although it seems to be a live question to what extent unconsciousness is part of implicitness, it is perennial in the discourse and should be mentioned when defining implicitness in this context. I will, following Brownstein, add a clause which will serve the purpose of acknowledging the debate in my own definition.

As seen in section 1.1.1, I mentioned not only "implicit bias" but also "implicit cognition" and "implicit attitudes". The two latter are often used in the psychological discourse to describe a wider field than IB. Psychologists who work in the field of "implicit social cognition" often study implicit attitudes toward different topics like political values, food, alcohol and drugs. Commonly, IBs refer to a subset of implicit attitudes, a subset which denotes implicit attitudes toward socially stigmatized groups such as women, blacks or homosexuals (Brownstein, forthcoming). I fail to understand why implicit biases only refer to our implicit (in the sense described above) attitudes towards stigmatized groups and not all social groups. The process does not seem to me as a different one simply based on which group is targeted. If I associate woman with home and man with work, these would both be biases which both can have negative consequences. Women might not get hired, and men might fear disconfirming the stereotype by not staying home with their children etc. I will use IB to refer to all implicit attitudes toward social groups.

To summarize, implicit bias is a subset of implicit attitudes. For now, it is sufficient to conclude that they are, to large extent, automatic and unconscious tendencies to associate certain traits with members of particular groups. I will later on add to, and subtract from, this definition but for now it constitutes a good starting point for my investigation.

1.2 The Tests

I have in the previous section described IB on a very general level. In this section I will present the most common test (the IAT) for identifying implicit associations and some of the supposed errors we make by having such associations. When discussing the errors, I use the philosopher Jennifer Saul's description. The purpose of this section is to provide some

practical understanding of the concept and by examining some examples, further develop my definition of IB.

1.2.1 Implicit Association Test (IAT)

The IAT is a test designed by psychologists to "measure[...] the strength of associations between concepts (e.g., black people, gay people) and evaluations (e.g., good, bad) or stereotypes (e.g., athletic, clumsy)". (About the IAT, 2011) Daniel Kelly and Erica Roedder, in their article "Racial Cognition and the Ethics of Implicit Bias" write about the IAT and describe their method for understanding the test. They say: "Rather than provide a technical description of how the test works, it will be more useful to convey its flavor." (Kelly & Roeddert, 2008, p.524) I to will use their method to convey the main idea of the IAT. Consider the following scenario: You are asked to sort words from the list below as quickly as possible, sorting the positive adjectives and Black names in column A, and negative adjectives and White names in column B.

Lakisha, Delicious, Sad, Jamal, Peter, Death, Happy, Unhappy.

Now, consider that you are asked to reiterate the test but this time sort the positive adjectives with the white names in the column A and the negative adjectives with black names in column B. What is discovered is that, in most cases, you would find it easier to sort the words when the positive adjectives were paired with white names and negative adjectives with black names.² In this case "easy" refer to faster and with less error. Kelly and Roeddert write:

The core idea behind both our [...] exercise and actual IATs is that stronger associations between items will allow them to be grouped together more quickly and easily (Kelly & Roeddert, 2008, p.525)

The main idea behind the IAT is that faster and more accurate performance in pairing positive adjectives to white names than positive adjectives to black names indicates stronger association between good and white. In turn, Kelly and Roeddert add that stronger associations between good and white "are taken to indicate an implicit bias towards"(Kelly & Roeddert, 2008, p.525) whites over blacks. The IAT is conducted on computers and might

² The "you" referred to is a generic test-taker. The current reader of this paper might not represent such biases (but you most probably do).

differ in a few ways from just described, but for the purpose of this paper, it is close enough to explain the processes at work.³

The IAT-results fulfill both the automaticity criterion (hard to control, non-attention demanding, unlimited capacity) and the unconsciousness criterion (difficult to locate through introspection). But what do we learn from the IAT? Is it a problem that we pair good with white or bad with black? To consider the implications of having an implicit bias, I will regard a number of problem areas which reveal some of the issues.

1.2.2 The CV-error

The CV-test is a very clean and simple test. The subject is asked to rate a number of CVs and within their stack of CVs there are doubles, i.e. identical CVs except for a small difference: the name on the top of the CV. One test show that we are more likely to rate a CV with a Male-sounding name better than the same CV with an Female-sounding name. Variations of the experiment, i.e. changing the name to test other underprivileged social groups, often show similar results. In these tests, the subjects have been asked a number of questions, Saul writes:

The right name makes the reader rate one as more likely to be interviewed, more likely to be hired, likely to be offered more money, and a better prospect for mentoring. (Saul, 2012, p.244)

The above quote is a bit misleading since it states that the reader rates some persons with the right name as "more likely to be interviewed[...]" than some others. This would not be surprising if the test subject was familiar with the theories of implicit bias, believes in any kind of structural discrimination or has read any statistics on the topic. It would simply be true – a woman is less likely to get interviewed, hired etc. What Saul actually suggests, I believe, is that the test subjects were more likely to rate specific instances of the CV higher. For example, the test subject is more likely to consider the person behind the CV with the right name as more hireable.⁴ Saul continues:

³ It is to be noted that this test is not, in psychology, regarded as entirely uncontroversial. In Hart Blanton and James Jaccard's article "Arbitrary Metrics in Psychology", they suggest that the IAT-results are often arbitrary. Blanton and Jaccard's view has, in turn, been criticized by Greewald, Nosek and Sriram ; claiming that the IAT has mechanisms for dealing with said arbitrariness.

⁴ Support for this interpretation can be found in Saul's suggested further reading. Namely: "Steinpreis, R., Anders, K., and Ritzke, D. 1999. The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study. Sex Roles, 41: 7/8, 509–528". The article's results are of the type I am suggesting above.

Moreover, the person making these mistaken judgments is surely unaware of the role that social category is playing in the formation of their views of the candidates. (Saul, 2012, p.244)

The test subjects are likely to be unaware of the impact their presupposed views on social category has on their judgment of CVs.

We can quite clearly see real implications of the biases. We are making decisions which affect people's lives significantly, based on factors which many of us would judge to be irrelevant – the social category. The ideal hiring situation would be one where the recruiter rate CVs after a given set of standards. These standards should be the same, indifferent of which CV is rated. When confronted with the fact above, that names make a difference, we must either conclude that the standards have changed or the judgment of the CV has changed. To my investigation, it is not important to decide which, but it is worth noting when related to a name, something happens with the judgment of the CV. This is the error I want to highlight. It simply seems unfair – the name should not matter if the content is the same!

1.2.3 Prestige Bias

Another type of error which can be caused by having implicit attitudes is the so called prestige bias. In this test, previously published papers were reemitted to psychology journals but with two things different, the names and the affiliations were less prestigious. What was astounding was the fact that only 8% of the journals detected that the papers had already been submitted. 89% did reject the papers, but not on the grounds expected (like plagiarism), they were rejected on grounds of serious methodological errors. What to make of this? Saul says:

These are experts in their field, making judgments about their area of expertise — psychological methodology — and yet they are making dramatically different judgments depending on the social group to which authors belong (member of prestigious VS non-prestigious psychology department). (Saul, 2012, p.245)

Everyone makes mistakes, but this is kind of ideal circumstances. They should not be making these kinds of mistakes but nevertheless they do. This type of example gives us a wider playing field when investigating the implicit bias. If this is an instance of

implicit bias, which I very much believe, then we can use "temporary" social group-belongings like work or social class and not only "fixed" like gender or ethnicity.⁵

1.2.4 Shooter Bias

The last type of bias-test I will describe is one commonly referred to as the shooter bias or weapon bias. This is a test in which the subject is asked to identify some ambiguous object in the hand of a person. Saul writes:

In these studies, it has been shown that the very same ambiguous object is far more likely to be perceived as a gun when held by a young black man and something innocent (like a phone) when held like by a young white man (Saul, 2012, p.245)

She continues:

These [studies] show that implicit bias is getting to us even before we get to the point of reflecting upon the world[...] (Saul, 2012, p.245)

A psychologist who formed such a test, B. Keith Payne, summarizes the importance of such a finding quite well. He claims that when officers of the law or other authorities are forced to make split-second decisions, these associations could be a matter of life or death. (Payne, 2006, p.287) In America, we have this year seen devastating examples where split second decisions by police officers has lead to horrible mistakes. One of the latest concerns a 12-year-old African American child who was "[...] fatally shot by a Cleveland police officer over a toy gun" (Meagan Keneally and James Hill, 2014). We do not know if this is due to implicit biases, although, if it were, the scenario would not differ greatly from the situation above.

1.2.5 Conclusions so Far

In these errors, we can observe that the ideal situation is where all individuals get judged the same way when exhibiting the same set of (relevant) properties. Unfortunately, something happens on the way. We are influenced by our biases against certain social groups; may it be the bias male over female, the prestigious over non-prestigious or the white over black. Worth noting is the fact that the test-subjects in the

⁵ A temporary social group is one you can leave at will: you do not have to work for a specific department. It is less easy to leave a fixed social group.

error cases not only form judgments based on implicit biases but also change their line of actions. They publish, hire and avoid shooting the person with the correct affiliation, name and looks.

Now, allow me to iterate my first definition of IB. Implicit biases are largely automatic and unconscious tendencies to associate certain traits with members of particular groups. After observing our errors, I would like to add a functional clause, namely that implicit biases must also affect our judgments and/or actions. If IBs did not influence our behavior or judgments, IBs would achieve significantly less attention. Rather, it is their effect on judgment and action that make IBs philosophically and psychologically interesting.

I will later discuss the fact that not all implicit biases have the same affect on guiding our actions, but I will claim that it is of no major importance to my paper — more of this later on. For now, we can produce a definition of implicit bias that is sufficient for my purpose:

Implicit biases are, to a large extent, automatic and unconscious tendencies to associate certain traits with members of particular groups in such a way that affect judgment and/or interaction with members of the particular groups.

1.3 Preempting Criticism

One way in which we can define implicit bias is through wide functional definitions like mine above. This is not uncommon in the discourse. The earlier mentioned Jennifer Saul, who is one of the leading philosophers in the field of implicit bias, writes that implicit biases are "unconscious biases that affect the way we perceive, evaluate, or interact with people from groups that our biases 'target'". Both mine and Saul's definitions include a functional clause which states that the implicit bias must affect our actions or judgments. To exemplify let us bring back Bob. Imagine that Bob takes the IAT and the test reveals that he has strong associations between negative adjectives and Arabs. To both mine and Saul's definitions, this is not enough to be the bias we are after; Bob's actions or judgments must also be affected by his associations.

This type of definition has been confronted with serious criticism by the philosopher Jules Holroyd and the psychologist Joseph Sweetman. In their article "The heterogeneity of Implicit Bias", they claim that when defining IB in a broad, functional sense (like I do), IBs cover a heterogeneous set of psychological processes which in some cases proves problematic. One of the ways in which my type of definition covers a heterogeneous set of processes is that different biases are related to behavior in different ways. If Bob has IB1 and IB2, IB1 might influence his behavior in a different way than IB2 does (this is supported by psychological empirical evidence). A second way is that our explicit beliefs affect our IBs in different ways. Consider again Bob's IB1 and IB2. IB1 might be influenced by his explicit beliefs when IB2 is not (again this is supported by empirical evidence). Now this does imply that I cover a set of heterogeneous processes, but does it matter? Holroyd and Sweetman suggest two main problem areas which I choose to name metaphysical and normative. These two areas can be summarized by the following quote:

Gaining an accurate understanding of the problems, and being in a position to make specific and useful normative recommendations, is hindered by the use of the term implicit bias in this broad way which does not specify what kind of associations are at issue. (Holroyd & Sweetman, forthcoming, pp12-13)

The metaphysical problem is constituted by the idea that without a precise understanding of the IB, we cannot know what kind of problems spring from it. The normative problem suggests that the wide definitions might cause us to give the wrong type of advice when aiming at militating the negative effects of IB. If Bobs IB1 can be militated by preemptive measure A, it is not certain that Bobs IB2 also can be militated by A. This might lead us to give the wrong normative recommendations.

Although this seems true, it does not affect my paper significantly — I am not looking to make any normative recommendations. My aim is to identify when and why an IB is ethically problematic and when it is not — the question is not how to militate but rather when and why we are to militate. As discussed in **1.2.5**, it is because of the effects IBs have on action and judgment that they are of philosophical and psychological interest. The ethically problematic situations are exclusively constituted by situations where our IBs affect our behavior. Reconsider the case with Bob where we identified that he has strong associations between negative adjectives and Arabs. If the associations are not affecting him in any practical way and he is explicitly against any form of discrimination, it would be wrong to say that his

having this bias is wrong. I would further argue that the fact that there is a difference between IB1 and IB2 when considering how they affect our behavior and how they are related to explicit beliefs is of little or no importance. I am interested in the IBs that produce an ethical problem — the ones that do not, are of no interest for my current investigation. My wide functional definition allows all implicit associations that influence our actions or judgments. It does not matter if they vary in strength. To more precisely define IB would be superfluous work, what I need my definition to capture is already captured by it.

2 What is Wrong with Implicit Bias?

Consider the following case: Our old friend Bob is looking to hire a new assistant. In accordance with earlier he is implicitly biased against Arabs and this causes him to act in certain ways (not hiring equally qualified Arab applicants etc.). When confronted with this type of scenario, most of us would say that something is wrong. What constitutes the wrong is less clear. The fact that Bob is unaware of, and not in control of his actions, does not seem to constitute what is morally problematic. Imagine a case where Ken, Bob's counterpart at a rival philosophy journal, deliberately and consciously chooses to not hire equally qualified Arab applicants. In this case, Ken seems worse than Bob, even though Bob is the one with the implicit bias. I suggest that awareness and control from the agent's point of view is a question of moral responsibility: can one be held responsible when one is not aware and in control? This is indeed an interesting question but it is not for me to debate in this paper.

I suspect that our (at least my) intuition that something is morally wrong in Bob's situation hinges on a conception of discrimination and the wrong thereof. What seems to be happening (and wrong) in Bob's case is a division between people, where on one side we have Arabs (who get treated worse) and, assuming he has no other biases, everyone else on the other side (who gets treated better). This way of dividing people seems to be very similar to pre-philosophical understanding of discrimination. To examine if the wrong of certain IBs can be explained by reference to discrimination, we must first examine what discrimination actually is and, further, what is wrong with it.

In the following sections of this chapter, I will attempt to develop a definition of discrimination, discuss what might be wrong with discrimination and finally, I will attempt to merge the discussion on discrimination with my previous discussion on implicit bias.

2.1 What is Discrimination?

To understand what is meant by discrimination it might be enlightening to answer the question, 'what is being done to whom and why?'. Initially, to answer the question of 'whom?', it seems as if we are dealing with people who are perceived as members in certain social groups. Can it be any social group? To answer this subsequent question we might get some

help from Kasper Lippert-Rasmussen. Lippert-Rasmussen claims that the social group cannot be just any social group; it must be a socially salient one. According to Lippert-Rasmussen "[a] group is socially salient if perceived membership of it is important to the structure of social interactions across a wide range of social contexts." (Lippert-Rasmussen, 2013, p.30)

Lippert-Rasmussen argues that social salience influences the structure of social interactions and it does so in a wide range of contexts. To exemplify something that is not social salience, consider the group of people who were born in towns named after kings. This would be a clear example of something non-socially salient: the fact that I am perceived as someone born in a town named after a king will hardly ever influence people's interactions with me. By contrast, my perceived sex, race or religion will influence social interactions in a wide range of contexts. Social salience is a matter of degrees and it is impossible to name an absolute number of contexts in which the social group must affect the interactions. For our purposes, it is enough to grasp the two extreme points of very low degree of salience (as in the town named after a king) to very high degree of salience (sex, religion and ethnicity). (Lippert-Rasmussen, 2013, pp.30-31)

Why should we add social salience? According to Lippert-Rasmussen, we win a lot by speaking of social salient groups. The first point is in line with his later theory on what is wrong with discrimination. Lippert-Rasmussen claims that cases where discrimination is targeting, e.g., people coming from towns named after kings, are idiosyncratic. The person discriminating is doing it on a "whim" and Lippert-Rasmussen continues: "[i]t is just that in the great majority of cases they will not seriously harm the disadvantaged party, precisely because of their idiosyncratic nature." (Lippert-Rasmussen, 2006, p.169) He continues: "This is not to deny that such idiosyncrasies can be as bad as, and reflect as corrupted a character as, genuinely discriminatory acts." (Lippert-Rasmussen, 2006, p.169)

The second point is, social salience can explain why we do not speak of discrimination when differentiating between the qualified and the unqualified, the deserving and the undeserving etc. The class of all people not being qualified for any job has no members and neither does the class of all people undeserving of everything. This might seem like an oversimplification by Lippert-Rasmussen. To be qualified and deserving seems to be a central thing in our society; there seems to be a thick line drawn between the qualified and the unqualified. What I believe Lippert-Rasmussen would have responded to this objection is that the salient group also needs stability. Although it is true that most of us are unqualified for most jobs, there are

always some of us that are qualified. The group would vary too much to even be called a group, and the stable group (of those not qualified for anything) would be membered by none. I will recur to the question of why social salience is important to discrimination after further discussion in later chapters.

Let us return to the 'what is being done' -part, which will also pour over to the 'why' part. In everyday language, discrimination is something which targets persons, and based on their perceived membership in social groups, they receive worse treatment than they would, had they not been perceived as members of that certain group. This is a good starting point in answering the 'what' and the 'why', but it has some shortcomings in philosophical sophistication. I suggest the following, more developed, version of our everyday conception of discrimination: discrimination is when persons perceived as members of a social salient group, by act, policy or law, are treated disadvantageously based on their membership in said group. First noted is that one cannot be positively discriminated against (towards). Consider the following situation: person x gets treated better than person y based on x's membership in the socially salient group S. It is not "positive" discrimination for x; rather: relative to x's treatment: y is getting disadvantages based on y's belonging, (or not belonging) to socially salient groups, e.g. a club only allowing white males is discriminatory against everyone not belonging to that social category. Using it differently seems too far off from how we commonly use the term, and, frankly, I can't see the advantages in diverging from the common use.

A second important point about our first definition is that the disadvantageous treatment is measured relatively and not absolutely. Consider the following scenario: a school class is on a fieldtrip. After traveling by bus and walking around in the forest all day, the children are very hungry. Now, consider that the teacher gives the black children half the amount of food that s/he gives to the other children. The black children would be better off than before, i.e. it was advantageous for them on an absolute scale, but they would still be discriminated against. This is due to the relative nature of discrimination, i.e. the disadvantage imposed on the discriminated is relative to how others are treated.

The persons referred to in the definition above include both individuals and groups. What is meant to be captured by 'law, policy or act' is some of the ways in which one can be discriminated against. My three 'ways' are very similar to what is commonly used in the main

discourse, e.g., in Andrew Altman's Stanford article, he claims that "acts, practices, or policies" (Altman, 2011) are what is discriminatory. I believe that Altman's use of "practices" cover more ground than my use of "law" does. I will therefore exchange "law" in my above description with "practices". With that said, I believe we can respond to the 'what is done to whom and why' in a satisfactory way.

What is done: an act, policy or practice imposes a relative disadvantage. **To whom:** Individuals and groups who are perceived as members of, or constitute, a socially salient group. **Why:** on the basis that they in fact are part of said socially salient group.

From this we can create the following definition, let us call it **D1**: x is discriminating iff x is an act, policy or practice which imposes a relative disadvantage on individuals or groups who are perceived as members of, or constitute, socially salient groups on the basis that they are perceived as members of , or constitute, said socially salient group.

2.2 What Makes Discrimination Wrong?

We have got this far mentioning nothing of what makes discriminatory acts wrong; now, let us face the morality of the concept. To do this I will borrow two notions from Altman's Stanford article: discrimination as a moralized concept and discrimination as a non-moralized concept. The moralized concept says that discrimination is a thick normative concept. A thick normative concept is a concept, like bravery or cruelty, which has descriptive content as well as evaluative content. To say that someone is brave is to praise that person and conversely, to say that someone is cruel is to criticize that person. To say 'discrimination is wrong', according to this view, is a tautology. The non-moralized concept claims that discrimination is only a descriptive notion and, if we say that discrimination is wrong, we say something substantial. To further describe the two positions we, could use D1 to exemplify. D1 seems to say nothing of the wrongness or rightness of discrimination and seems to constitute a version of a non-moralized conception of discrimination. If one wants to use the moralized conception something must be added. One could simply claim that the disadvantage which is imposed onto the person being discriminated against is wrongfully imposed. Using the evaluative notion of wrongfulness adds what is needed for the concept to become a thick normative one. (Altman, 2011)

Most positions allow both the non-moralized and the moralized conceptions of discrimination and chose to focus on the latter since it hosts an interesting moral question, i.e. 'what is wrong with discrimination?'. I will in this paper present two views: the harm-based view and the dignity-based view. These views are in no way exhaustive of possible ways to answer the question⁶ but they represent two strong positions in the discourse. In short, the harm-based view claims that discrimination is bad when the discriminated is harmed by it, the dignity-based view claim that the act of discrimination is wrong per se since it does not acknowledge the discriminated as a fully moral person.

2.2.1 The Dignity-based View

Sometimes, making distinctions between people is morally permissible, e.g., it is acceptable to discriminate against people under 18 by saying that they are not allowed to drive, and sometimes it is impermissible, e.g., it is unacceptable to discriminate against women in a hiring process. To decide when such distinctions are permissible or not is to solve what Deborah Hellman calls the discrimination puzzle. In her book "When Is Discrimination Wrong?" (2008), Hellman attempts to solve the puzzle.

As suggested by the formulation of the discrimination puzzle, Hellman adopts the distinction between the moralized and the non-moralized conception of discrimination, she also seems to accept their coexistence. Consider two different social groups: our earlier mentioned group of people coming from a town named after a king and women. Now, consider a first scenario (scenario A): a university that has as an admission policy that no one coming from a town named after a king is to be admitted to the university. Consider now our second scenario where, the group is constituted by women (scenario B): the same university added an additional criteria stating that no women are allowed. Hellman suggests that scenario A is discriminatory and wrong but it is not wrong on the basis of discrimination; it is not the right kind of wrongful discrimination. Scenario B, on the contrary, does constitute the right kind of wrongful discrimination.

⁶For further reading on the different alternatives, Altman present a comprehensive summary in his Stanford Encyclopedia entry: Altman, Andrew, "Discrimination", *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition)

To explain the above distinction we need to back up a bit. Hellman's endeavor of solving the puzzle starts at a "bedrock moral principle" which is the equal moral worth of all persons. Let us, for simplicity, name that principle the principle of equal dignity⁷. Hellman claims that the principle of equal dignity is constituted by two parts:

1. There is a worth or dignity of persons that demands that we treat each other with respect. What constitutes respect is a different question for another time.
2. The dignity or worth presented above does not vary with other traits. If x has a higher level of intelligence, kindness or any other trait than y, this does not imply that x has more moral worth or dignity.

Hellman suggests that the moral concern we have with drawing distinctions between people is that in doing so, we might act in ways that fail to treat others as equally worthy. The question of when discrimination is wrong can then be translated into when "[...] does drawing distinctions among people fail to treat those affected as persons of moral worth?" (Hellman, 2008, p.7)

Let us now try to make sense of the previous distinction between the right and the wrong kind of wrongful discrimination. As we saw earlier, Hellman's suggestion of what constitutes wrongful discrimination is where someone gets treated in ways that fail to recognize his or her inherent moral dignity or worth. According to Hellman, scenario A is not such a situation. Hellman suggests that our intuition, that something is wrong with this specific criteria of admission, is based on other things than the principle of equal dignity. It might be the fact that our intuition regarding what a university is supposed to be is getting frustrated. We might want a university to be a haven for meritocracy⁸ and said criteria might prohibit such a haven at this university. The point is that when people coming from towns named after kings are discriminated against (in the manner above), it is not the correct form of wrongful discrimination.

Scenario B does fail to recognize the inherent moral worth or dignity of the persons discriminated against. What is different in these cases? Hellman says: "Being denied a job on

⁷ Much can, and has been said on the topic of the equal dignity. In current debate on equality, the standard interpretation of the principle of equal dignity follow in the footsteps of Dworkin. Dworkin claims that moral equality should be understood as a prescribing treatment of persons as equals. This suggests not that everyone should be treated equally (the same) but they should be treated with equal concern and respect. (Dworkin 1977)

⁸ Hellman contests the claim that what is wrong with discrimination is the fact that we are not being distinguished based on merit. She does so arguing that it is very difficult to tell when merit counts or not e.g. if someone is very merited but a super-nazi then merit wouldn't count. For further reading see *When is discrimination wrong*, chapter 4 by Deborah Hellman.

the basis of being female demeans women in a way that being denied a job because one's [coming from a town named after a king] does not." (Hellman, 2008, p15) One might notice that Hellman writes 'demeans', which has not been discussed earlier in my paper. To demean someone is to not respect the moral worth of that person and, therefore, demeaning is what is wrong with discrimination. To understand the difference between scenario A and scenario B, the natural question to follow is: when is someone being demeaned?

What can be distinguished between the two groups in scenario A and B is the fact that the group of females has a significant history of being assigned lower status which the group from the A scenario lacks. The historical and current social status seems to matter in some way in the current puzzle and Hellman offers one way of explaining why this is: when people are separated on the basis of some trait, in certain contexts the people separated might be demeaned; i.e., the separation is wrong. An utterance is demeaning or not based on the meaning conveyed by the utterance. Meaning can come from three sources: 1: The intent of the speaker 2: The interpretation of the listener, and 3: The context in which the utterance is made.

Hellman suggests that the intent of the speaker cannot be the only way in which one can be demeaning. Consider a teacher who commands his or her students to sit sorted by race with the intent that s/he simply likes the aesthetics of uniformity. Considering the history of racial segregation this might, even though the intent seems harmless, demean some of the persons being ordered to sit in this manner. What is demeaning according to Hellman is the context rather than the act. This demeaning act often comes with a side effect. Generally when people are demeaned, they feel treated badly and might suffer psychological harms. This effect is not what Hellman wishes to emphasize. The fact that the teacher order his or her students in this manner is demeaning and wrong by itself, not by virtue of the effects produced by this order. The denoted significance of historical and current status seems to support the idea of social salience discussed in section **2.1** The non-socially salient group, i.e. the group that does not figure in many social contexts or influence the structure of social behavior, is very unlikely to have any significant status which can influence our context.

The discussion regarding the possibility for all three ways of conveying meaning to be demeaning in the correct sense, I will leave to others⁹. In my attempt to explain what is wrong with IB, it is enough to say that Hellman's description of what is wrong with discrimination can deal with actions caused by IBs. It is possible for the context to make an unintentionally demeaning act demeaning and, therefore, discriminatory wrong. In situations where the context does make the act demeaning, we can explain our intuitions about the wrongness of IB. Our IBs constitute discrimination of the wrongful kind. Even when the act is demeaning in the second way, the interpreter feels demeaned, it can explain the wrongness of IB. It is only the intentional way of being demeaning which cannot be used to explain the wrongness of IB, when it is wrong – but according to Hellman, this is the least important.

2.2.2 The Harm-based View

So far I have only briefly sketched the harm-based view as the view that claims discrimination as bad when the discriminatee is harmed by it. To fully understand what this could mean, we must elaborate. The harm-based view I am presenting is originally offered by the earlier mentioned Lippert-Rasmussen in his article "The Badness of Discrimination"¹⁰. Lippert-Rasmussen too, in his article, illuminates the distinction between the moralized and non-moralized conception of discrimination which he name the P-ist and the P-based discrimination. P-based discrimination, "[...]which involves treating individuals differently depending on their P-properties, but is not necessarily morally objectionable"(Lippert-Rasmussen, 2006, p.186) is the same as a non-moralized conception; P-ist, "[...]which involves treating individuals differently on the basis of their P-properties in a morally objectionable way" (Ibid) is coextensive with the moralized conception. When investigating the *badness* of discrimination, Lippert-Rasmussen attempts to: explain why P-ist discrimination is wrong and describe when and why P-based discrimination is wrong.

Initially Lippert-Rasmussen claims that discrimination, when it is bad, is bad pro tanto. To say that something is pro tanto bad is, according to Lippert-Rasmussen, to say that "[...] from a moral point of view [it] is in one respect regrettable" (Lippert-Rasmussen, p174). He does not claim that it is prohibited, or, all things considered, forbidden, only that it is in one aspect

⁹ For further reading see Hellman, 2008 chapter 2-3

¹⁰ And further developed in his book "Born Free and Equal?: A Philosophical Inquiry Into the Nature of Discrimination"

regrettable. When can we say that discrimination is pro tanto bad? According to Lippert-Rasmussen:

an instance of discrimination is [pro tanto] wrong, when it is, because it makes people worse off, i.e., they are worse off given the presence of discrimination than they would have been in some suitable alternative situation in which the relevant instance of discrimination had not taken place.
(Lippert-Rasmussen, 2006, p.155)

This formulation might prima facie seem trivial; whatever the process, it would be pro tanto bad if it caused someone to be worse off. Although this might be the case, Lippert-Rasmussen claims that his position is important in the current discussion on discrimination. First, he claims that in the current discourse, discrimination is almost always referred to as a thick normative concept— the above formulation allows discrimination to be contingently bad.¹¹ Second, if someone did accept that discrimination is contingently bad, few would agree that the badness lies in the harms it causes. Saying that discrimination is bad based on the harms done is to bring something new to the table.

The harm-based view Lippert-Rasmussen offers here is not a finished package; there are a number of possible positions to pick from within the harm based view. Distinguishing between positions is a way of understanding the harm-based view better. The first distinction made is the question of a relevant baseline for determining whether a person has been made worse off. Lippert-Rasmussen suggests that we here have two options, the first being: discrimination is bad when it makes the discriminatee worse off than she would have been had she not been subjected to it. The second option is what he calls the moralized baseline, which claim that discrimination is bad when it makes the discriminatee worse off than she would have been in a just, or morally best, outcome. Using the moralized baseline, one can explain intuitions like why reverse discrimination might be permissible – this would be because both parties seem to get closer to the just outcome while non-reverse discrimination tends to widen the gap. When using the moralized baseline, one needs to decide what a just outcome is, but this is a task for those looking to defend the harm-based view from a

¹¹ I would have to disagree with Lippert-Rasmussen on that point. Hellman in her account, which Lippert-Rasmussen offers critique later in his article, does accept a non-moralized as well as a moralized conception of discrimination. Further the earlier mentioned Stanford-article on the topic also acknowledges the distinction but focuses on the P-ist or the moralized kind of discrimination. I suppose that the discourse Lippert-Rasmussen mentions is not the philosophical but rather the legal discourse which has more of a thick normative concept - account of discrimination.

moralized baseline and not of any interest for my paper. Additional ways of distinguishing between harm-based positions is to investigate how we determine how well off a person is. It may vary in a number of ways: if the worse off is regarding one's entire lifespan or if it is regarding certain life segments; if the worse off is regarding individuals or groups and so on and so forth. One thing we can acknowledge right of the bat regarding the harm-based view is that it is not dependent on intentional actions. An action would be discriminatory even if it was based on an IB.

2.2.3 Implicit Bias and Discrimination

I have in the two above chapters discussed two strong positions in explaining what is wrong with discrimination. I have argued that both views can explain what is wrong with implicit bias, and by showing this, I prove that it is possible for IB to be bad in virtue of it being wrongful discrimination. It is possible but it in no way proves that it is the case. To view this possibility in a more favorable light, we need to argue further.

IB is a process which causes us to make judgments or act in different ways in relation to people, based on them belonging to groups. This might not always be problematic but when it is, it seems that the wrong lies within the distinguishing between groups: the correct or wrongful distinction between groups is what constitutes the aforementioned discrimination puzzle and, therefore, the two debates coincide.

Consider the following scenarios:

1. A position at Bob's philosophy journal is vacant and Bob is looking to hire the most capable person for the job. Implicitly, Bob associates women with home, which affects how he judges women's capacity for work and ends up hiring Stephen, who was less qualified than Karen (due to Bob's IB).
2. Annie gets off the train in a town unknown to her. She implicitly associates people working at 7-Eleven with knowing their way around town better than people working at Pressbyrån (a similar establishment). This causes her to ask for directions at the adjacent 7-Eleven instead of the equally adjacent Pressbyrån.
3. A university conducts an interview as a part of their admittance procedure. One of the interviewers, Kaylee, has an IB against people coming from towns named after kings. Her IB causes her to choose other applicants, were such a person to apply.

These examples of IBs seem to vary in wrongness. 1 and 3 seem prima facie bad where 2 seems non-problematic. As I mentioned above, my suggestion is that the wrongness of IB, when it is wrong, hinges on the wrongness of discrimination. Allow me to iterate my definition of discrimination.

D1: x is discrimination iff x is an act, policy or practice which imposes a relative disadvantage on individuals or groups who are perceived as members of, or constitute, socially salient groups on the basis that they are perceived as members of, or constitute, said socially salient group.

What we can observe from this definition, which takes no part in when a discriminatory act is wrong, is that scenario 2 seems to be excluded. The group of people working at Pressbyrån does not seem to constitute a socially salient group. As seen in **2.1**, a socially salient group must influence the structure of social interactions in a wide range of contexts, which said group does not seem to do. Further we can observe Hellman's point regarding the importance of historical and current social context. The group of people working at Pressbyrån have no history of being oppressed neither a current lower status than others. Hellman's analysis would claim that it would not be demeaning for the people working at Pressbyrån, therefore, it would not be wrong. If we instead use Lippert-Rasmussens analysis, we could say that the persons at Pressbyrån are not worse off due to the action and therefore not wrong¹². Using wrongful discrimination to determine what is wrong with scenario 2 seems to match our intuitions on the subject.

Consider now scenario 1. Bob is back and this time he is biased against women. This case seems to constitute discrimination according to D1 and it seems to be wrong. Again, starting at Hellman's analysis, we can conclude that the group of people (women) has a history of being oppressed and still suffer from different forms of oppression – lower salaries, less frequenting positions of power etc. To base this kind of decision on women's belonging to the social group "women", even if the decision is not controlled or conscious, is morally

¹² One advantage with Lippert-Rasmussens analysis is that it also can capture a further intuition, namely: does it matter if one chooses to ask for directions at one place or the other? The seriousness of the topic seems to influence the wrongness of the act. It matters more if the thing at stake is the wellbeing of a person than if the same person is asked or not asked for direction. I believe that Hellman's analysis would acknowledge this distinction by saying that frustrating the wellbeing of a person due to discrimination is not respecting and therefore to demean the person.

impermissible. Considering Lippert-Rasmussens analysis, we would get the same answer – Karen is worse off due to the distinction and this would hold even if it was produced by an IB.

Finally, observe scenario 3, which is borrowed from the above discussion on Hellman's notion of discrimination. What we see here is a group which most certainly is not socially salient, but the division is most certainly wrongful. This is the case which I named the wrong kind of wrongful discrimination. As we can see from the above discussion in **2.2.1**, according to Hellman, this distinction is wrongful but not wrongful discrimination. What is wrong, as suggested above, might be that we fail to accomplish what we want a university to be. The IB is wrongful in virtue of Kaylee frustrating the idea of a merit based university. As we saw in **3.1**, Lippert-Rasmussen too acknowledged the difference between discrimination containing socially salient groups and containing non-salient groups. Let us call the first salient discrimination and the second non-salient discrimination. Hellman and Lippert-Rasmussen both agree that the non-salient discrimination can also be bad, just not on the same grounds as the salient. I suggest that the concept of wrongful discrimination is heterogeneous and it covers both the salient and non-salient type of wrongful discrimination, which, in turn, denotes two different sets of moral problems.

To cover all problems which might arise from IBs, we must also cover idiosyncratic IBs, e.g., people coming from towns named after kings. To do this, we must accept a wide notion of wrongful discrimination which is constituted by two subsets: the first, our common use of wrongful discrimination, which claims that they regard salient groups. Hellman and Lippert-Rasmussen have in their separate ways suggested an answer to what is wrong with that particular kind of discrimination. It would be of philosophical interest to answer the further question of who is right, Hellman or Lippert-Rasmussen, but due to limited time and space, I will not be able to conduct such an investigation. The second subset is regarding non-salient wrongful discrimination which can be wrong in different ways, e.g., merits. I would like to emphasize that the second subset is much less important due to its limited scope. Our focus should lie on mitigating the effects of the salient wrongful discrimination because it has a massive influence in our lives, an influence the non-salient wrongful discrimination has not.

To further underline the difference between the salient and non-salient wrongful discrimination, we can investigate the IB-errors discussed in **1.2**; the CV-error, the prestige bias and the shooter bias. The first presentation of the CV-error consisted in comparing male with female applicants. Gender as we concluded earlier has high level of social salience.

Female applicants were imposed a relative disadvantage, seemingly based on their perceived membership of a social group. This is a typical case of salient wrongful discrimination and can be wrong in the way defined by Hellman or Lippert-Rasmussen above. Let us now turn to the case of shooter bias, which too seems to be a clear cut case of salient wrongful discrimination. In this case black men are imposed a relative disadvantage, and the disadvantage is based on their perceived membership of the group of black men. Both ethnicity and gender, as discussed above, are socially salient and therefore our case of shooter bias is the salient wrongful discrimination. What is wrong with such discrimination is the same as in the CV-error and can be offered by Hellman or Lippert-Rasmussen. Let us now investigating the prestige bias. It seems as if the low-prestige affiliations get a relative disadvantage in relation to the high-prestige affiliations. The low-prestige affiliation and high-prestige affiliation does not seem to be typical socially salient groups. As discussed in **2.1**, social salience is a matter of degrees rather than absolutes, and it may be true that the affiliation group is more salient than the earlier group of people coming from towns named after kings. This however does not change the fact that a prestigious affiliation is, most probably, not a socially salient group. What is wrong with the prestige bias is non-salient wrongful discrimination. As I argued above, our focus should lie on mitigating the IBs that lead to salient wrongful discrimination, which would exclude this case. I still claim that this is what we should do due to its wider scope. Even though the prestige bias is bad it affects only a narrow group of people when comparing to e.g. wrongful discrimination against women. The wrongness with such discrimination is a bit more intricate, since the prestige scenario is in some sense a middle ground. It is not salient nor is it an idiosyncrasy — it is somewhere in between. I would suggest that the wrongness in this situation is based on the earlier suggested idea of our intuitions regarding what a scientific journal is all about (merit).

Introducing non-salient wrongful discrimination serves the purpose of explaining the cases which, left unexplained, could possibly diminish the credibility of my argument. When we allow the second type of discrimination, it is a lot more viable to propose that what is morally questionable with wrongful IBs is wrongful discrimination. I have in this section shown that wrongful IBs can be explained in terms of wrongful discrimination by investigating a number of examples. I have also shown that IB is not a thick normative concept due to its contingent badness.

3 Summary

I have in this paper proposed a definition of IB which differs in a small but significant way from how IB is usually defined. According to my definition, IBs can regard all social groups. I do not believe there to be any significant difference in the process whether it is regarding socially stigmatized groups or any other group. With doing this, I allowed non-socially salient groups to be affected by IBs. To respond to my original question: ‘when and why is wrongful IB wrong?’, I leaned against a contemporary discussion on discrimination. What seems to be wrong with wrongful IBs is the division amongst people in a wrongful manner. It is common in the discussion regarding discrimination to divide between the right kind of wrongful discrimination and the wrong kind of discrimination and focus on the earlier. I have named these two types of wrongful discrimination salient and non-salient wrongful discrimination. By discussing Deborah Hellman’s dignity based view and Kasper Lippert-Rasmussen’s harm based view, I show two ways of explaining what is wrong with the right kind of wrongful discrimination (the salient wrongful discrimination). To account for the non-socially salient groups, which my definition of IB allows, I opened up for the second interpretation of wrongful discrimination, namely the non-salient. I also suggested that the salient wrongful discrimination is more important since, due to the nature of socially salient groups, it affects our lives in a significant way which non-salient wrongful discrimination does not. Further, I believe that my examples in 3.2.3 has provided some evidence for the belief that IBs are contingently bad.

To my understanding, the shortcoming of my paper is twofold. The first problem lies within the unfinished discussion on which analysis of wrongful discrimination is the correct one. If it was decided which way of discussing wrongful discrimination is the better one, we could achieve a greater level of accuracy when judging which effects of IB to mitigate. The second problem is that my discussion on non-salient wrongful discrimination gives little direction on how we are to decide when discrimination in this sense is wrongful. This is partly due to the lack of time and space and partly because the discourse surrounding discrimination often skips that part to go straight for the more important salient wrongful discrimination. All things considered, I believe it is a good thing having obvious black holes in one’s paper which are begging to be answered. It is a good starting point for further investigations and a good way

Jakob Olofsson
900416

FPRK01
Fall term

of potentially expanding a smaller subject from something interesting to something philosophically valuable.

4 References

About the IAT, (2011) available from: <https://implicit.harvard.edu/implicit/iatdetails.html>, [2014-11-12;1400]

Altman, Andrew, "Discrimination", (Spring 2011 Edition), *The Stanford Encyclopedia of Philosophy*. Available from:

<<http://plato.stanford.edu/archives/spr2011/entries/discrimination/>> [2014-12-01]

Blanton, Hart and Jaccard, James, (2006), "Arbitrary Metrics in Psychology", *American Psychologist*, Vol 61(1), pp. 27-41. Publisher: American Psychological Association [Journal Article]

Brownstein, Michael, "Implicit Bias", (Forthcoming), *The Stanford Encyclopedia of Philosophy*, DRAFT 8.29.14.

Dworkin, Ronald, (1977), *Taking Rights Seriously*, Cambridge: Harvard University press.

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

Greenwald, A G.; Nosek, Brian A.; Sriram, N, (2006), Consequential validity of the Implicit Association Test: Comment on Blanton and Jaccard, *American Psychologist*, Vol 61(1), Jan 2006, 56-61

Hellman , Deborah, (2008), *When is Discrimination Wrong?*, Harvard University Press

Holroyd, Jules and Sweetman, Joseph (forthcoming). The Heterogeneity of Implicit Bias. In Michael Brownstein & Jennifer Saul (eds.), *Implicit Bias and Philosophy*. OUP.

Kelly, Daniel and Roeddert, Erica, (2008), Racial Cognition and the Ethics of Implicit Bias, *Philosophy Compass* 3/3 (2008): 522–540, 10.1111/j.1747-9991.2008.00138.x

Keneally, Maegan and Hill, James, (2014) available at: <http://abcnews.go.com/US/cleveland-cops-recklessly-shot-boy-12-toy-gun/story?id=27402837>. [2014-12-15]

Lippert-Rasmussen, K. (2006). The badness of discrimination. *Ethical Theory & Moral Practice*, 9(2), 167–185.

Lippert-Rasmussen, K (2013). *Born Free and Equal?: A Philosophical Inquiry into the Nature of Discrimination*

Olson, Michael A and Fazio, Russell H, (2003) RELATIONS BETWEEN IMPLICIT MEASURES OF PREJUDICE: What Are We Measuring?

Payne, B Keith, (2006), *Weapon Bias Split-Second Decisions and Unintended Stereotyping*, Volume 15—Number 6 Copyright r Association for Psychological Science

Saul, Jennifer, (2012), *Scepticism and Implicit Bias*, Disputatio Lecture 2012, University of Sheffield, BIBLID [0873-626X (2013) 37; pp. 243-263]