"We Will Find the Black Man Who Did This"
Conceptual and Ethical Studies of
Discrimination in Criminal Justice

Frej Klem Thomsen

Department of Culture and Identity
Roskilde University
2011

"We Will Find the Black Man Who Did This"

Newsspeaker: "DNA evidence has freed a Panola, Georgia, man today, after 12 years in federal prison. James Marshall was sentenced to life without parole in 1996 for the mauling death of a state park hiker, Janet Kelly. Throughout his time in prison Mr. Marshall maintained his innocence, claiming his jaws were not powerful enough to have inflicted the 6-inch deep bites covering Ms. Kelly's body, and that he did not eat Ms. Kelly's entire left leg and lower torso as charged." [Cut to Janet's father] "They say after he killed Janet, he spent five minutes rubbing his hairy back up against a tree to scratch it – this man is sick." Newsspeaker: "The mauling was so vicious that Janet's funeral was closed-casket. Buried with her was the last photograph she took before she died: a self-portrait of her and two bear cubs. The Panola Police Department today apologized for the inconvenience they caused Marshall, and say they plan to reopen the case and quote: "Find the black man who did this." - The Onion News Network (http://www.theonion.com/video/dna-evidence-frees-black-man-convicted-of-bear-att,14323/)

"We Will Find the Black Man Who Did This" – Conceptual and Ethical Studies of Discrimination in Criminal Justice

Ph.d.-thesis

Research group for Criminal Justice Ethics

Philosophy & Science Studies

Department of Culture and Identity

Roskilde University

November 2011

Table of Contents

Behind the scenes of "We Will Find the Black Man Who Did This"

The best part of the process of learning is looking back on a period of – oh, I don't know, just to pick a number, say – three years, and realizing how much you have discovered and explored, not only of things that you did not know, but of things that you had no idea it was possible to know. The not-so-good part is that in true Socratic fashion the horizon of your knowledge expands faster than the area of your knowledge. For every idea or theory you become familiar with, you learn of two which you really ought to get round to taking a better look at one of these days. The *worst* part is that the realization of how much you have discovered, which was only very recently beyond even your horizon of knowledge, brings with it an uneasy feeling of what and how much is likely to lie beyond your horizon yet. The not so good part and the worst part of learning in combination can be pretty intimidating. And I can say confidently that I am more scared today than I was when I started this project a little less than three years ago. I like to think that this is a good thing.

In the summer of 2008, I prepared an application for the Danish Research School of Philosophy (PHIS), which combined my interests in cultural diversity and political philosophy with the research goals of the Criminal Justice Ethics group at the Department of Philosophy & Science Studies, Roskilde University. The result was an application focusing on the issue of discrimination in the criminal justice system, an issue that seemed to be under-explored and philosophically interesting. The

application nearly nosedived at take-off when PHIS approved my application and offered to submit it to the national Research Council for Culture and Communication (FKK) – the people who actually control the money – but on the condition that I work in several alterations and clarifications. This should not have been a problem, but doing so turned into an amateur Jason Bourne-thriller, involving short deadlines, finding a working internet-café in Tuscany, having a friend hack into my home computer to transfer files, employing a different friend as layout-editor, and explaining two days of hard work in the middle of a long-awaited vacation to an incredulous girlfriend. In late November it turned out that my application had been approved, which retroactively justified the effort.

Helmuth von Moltke famously quipped that "no plan survives first contact with the enemy." My original notion was to take what I expected to be a well-defined concept of discrimination and see how it would play out in the special context of the criminal justice system. The discovery that there was no such thing as a 'well-defined concept of discrimination' (with one major exception; see below) came as a rather jarring shock. Naturally, in preparing the project, I had found that there seemed to be some confusion on several issues, that many authors failed to provide adequate definitions at all, and that those that did disagreed vehemently on key points. (Nickel, 1972; Brest, 1976; Woodruff, 1976; Singer, 1978; McCrudden, 1985; Gardner, 1989; Campbell, 1991; Alexander, 1992; Cohen, 1994; Rutherglen, 1995; Nickel, 1998; Wasserman, 1998; Radcliffe Richards, 2000; Ezorsky, 2001;

Halldenius, 2005; Loury, 2005; Connolly, 2006; Vallentyne, 2006; Heinrichs, 2007) But I believe I expected that my impressions of confusion were mostly due to my personal deficiencies, both in finding the proper literature and in understanding the literature I had found. Although a modest part of me insists that even today I should not entirely rule out these two possibilities, I have had growing confidence that it is in fact the case that there exists no consensus on either how to define the concept or what is morally problematic about it (when something is).

And then there was Kasper's work. I knew Kasper Lippert-Rasmussen as part of the small circle of Danish philosophers working with ethics and political philosophy, and am ashamed to say that, perhaps due to a "no prophet is accepted in his hometown"-type of effect, I did not initially pay his work the attention that it warranted. And when at one point, shortly after beginning work on the project in the Spring of 2009, I reread his "The Badness of Discrimination" (Lippert-Rasmussen, 2006a), I had the distinct feeling that there was something brilliant going on there, and that if only the author had taken another 10 pages to unwrap the dense arguments, it would have made complete sense to me. (An authorial sin, I feel, which merits at most a brief stay in purgatory when compared to the special hell awaiting the all-too-common writer who, if only they had spent 20 *fewer* pages, would perhaps have made *some* sense...) It took me, in fact, the better part of a year to fully comprehend that not only was here a crisp, concise definition of discrimination, but that I was essentially persuaded by it.

The two successive surprises left me with a rather different task than I had first envisioned. I spent quite a lot of time exploring the concept of discrimination and the differing theories about what is and is not morally wrong with it. And my first writing turned out to be a critique of one particular condition which frequently appeared in definitions, and which I found (and find) implausible. Indeed, much of the rest of my work was to follow this path, and explore details of the conceptualisation of discrimination, which at least has the virtue of proving Whitehead's opinion that "all philosophy is footnotes to Plato" wrong, since clearly the conceptual clarifications I have attempted qualify rather as footnotes to Lippert-Rasmussen. (Lippert-Rasmussen, 2006b; 2007a; 2007b; 2007c; 2008; 2009; 2010; 2011)

My stay at the Oxford Uehiro Centre for Practical Ethics during the spring of 2010 served to infinitely broaden my philosophical horizon, and proved one of the most intense academic experiences of my life. Inevitably, my stay there was just long enough that I started feeling less of "a legal alien, a non-Englishman in Oxford" (to paraphrase CBE Gordon Sumner) before I had to return to Denmark. But encountering international academic philosophy in its full glory gave me perspective in a way I doubt anything else could have. Apart from the heady experience of seeing such luminaries as Joseph Raz, Amartya Sen, Ronald Dworkin, John Broome and Derek Parfit, I spent much of my time there investigating the wrongness of discrimination, particularly Kantian accounts (Darwall, 1977; Kennedy, 1997; Harris, 2002; Arneson, 2006; Darwall, 2006; Hellman, 2008; Glasgow, 2009; Moreau,

2010), and reading Parfit (Parfit, 1984; 1997; 2002); the latter of which I have to admit was on the whole both more fun and more worthwhile.

When during my stay I got the opportunity to provide a paper for a conference on racial profiling (which, ironically, was organized by my supervisor Jesper Ryberg and took place in Denmark), this gave me the first real shot at investigating discrimination in the criminal justice system. My analysis here turned out to require elements of decision theory, Bayesian probabilism, expected utility theory and behavioural economics (Kahneman and Tversky, 1974; 1979; Schauer, 2003; Harcourt, 2004; Peterson, 2009; Kahneman and Tversky, 2009a; Kahneman, Knetsch and Thaler, 2009; Kahneman and Tversky, 2009b) And although my grasp of these topics can charitably be called rudimentary, I found working with them to be one of the most eye-opening and rewarding lessons of the three years. As it turned out, my analysis of racial profiling also ended up deliberately avoiding the term discrimination, for the same reasons that my later work on equality before the law would. I came to the conclusion at this point, that little if anything that needed to be said about these issues could not be said without using the term 'discrimination', and that talking of them in terms of discrimination was often more likely to be confusing than helpful.

As conclusions go, for a project on discrimination this might be taken to be somewhat disappointing. But I think there are brighter sides: First, even if, as it seems to me, bringing the concept of discrimination into an issue rarely contributes to the discussion,

there are contexts in which it is most natural to talk about the problem in terms of discrimination. When, as in these situations, the concept of discrimination is already or unavoidably involved it is helpful if we have a clearer understanding of what exactly discrimination means. I hope that my work may, in some small and probably long-delayed way, help with this. Second, parts of the conclusions I have drawn have implications for more fundamental normative theory. Although hardly a telling argument against it, the fact e.g. that disrespect is as far as I can tell incapable of plausibly accounting for the wrongness of wrongful discrimination poses a problem for respect-based accounts of morality more generally. Similarly, although I have not attempted to do so, I believe that my arguments against the moral significance of equality before the law could be generalised to target deontic egalitarianism more broadly. Third, it was the thinking about discrimination which enabled me to analyse and discuss issues of racial profiling and equality before the law the way I have, even if while doing so it has seemed most useful to me to pull up the conceptual ladder after me. Fortunately, as it turned out, there was no need to remain silent about the problems at hand once I had done so.

Apart from the central research project, I managed to somehow get myself involved in a number of other projects during the three years. Meeting David Edmonds in Oxford, and hearing his highly recommendable "Philosophy Bites"-series prompted me to produce a similar series of interviews in Danish, "Filosofiske Godbidder", which idea I am extremely grateful to him for giving me his approval for, considering its blatantly copycat nature. Not

one to stop at just one form of copycatting, I collaborated with Jakob Holtermann in editing and producing a Danish volume in the "5 Questions"-series, with interviews from a broad range of the best contemporary Danish philosophers. The book did not attract quite the fame and fortune we were certain it was destined for, but the process was rewarding and (in large degree due to Jakob) enjoyable. Obviously, my Ph.d. also involved certain amounts of teaching, where the interdisciplinary policy of Roskilde University saw me supervise student-projects ranging from the anthropology of Confucian thinkers Mencius and Xunzi to an evaluation of the Danish Save the Children's failed summer camp for poor single-parent fathers. I confess that I did not always feel my core philosophical competencies were entirely adequate to the task, but on the whole teaching has proven to be a mutually agreeable experience – or so I believe. I was fortunate also to benefit from contact with several research networks, both the Criminal Justice Ethics group at Roskilde University, The New Network for Legal Philosophy at the Faculty of Law, Copenhagen University, the Centre for Studies of Equality and Multiculturalism, Copenhagen University, the Nordic Network for Political Theory and its impoverished twin the Nordic Network for Political Ethics. All of these allowed me frequent access to conferences and seminars, where the opportunities to present my work were both a useful motivating factor in getting work done and a source of rich feedback, just as the opportunities to encounter the cutting-edge work of others was inspirational and eye-opening.

Somehow, I managed to find time for all of this in a three-year schedule. Only to face an unfortunately all-too-reasonable argument that it would be preferable to finish three months ahead of time, a change of schedule which made the finishing run slightly more of a hectic sprint and slightly less of the leisurely-if-exhausting jog that I had envisioned. But all is well that ends well. And it has been, as they say, quite a ride.

Frej Klem Thomsen

February 1, 2009 – November 1, 2011

A brief history of discrimination in philosophy

Much of the early discussion of discrimination in the literature started *in medias res* so to speak. Thus: "Comparative injustice consists in arbitrary and invidious discrimination of one kind or another: a departure from the requisite form of equal treatment without good reason." (Feinberg, 1974, p.299) Discrimination was, it seems, taken to be a familiar phenomenon, and the concept did not seem in need of clarification or explication. Rather, like Feinberg, we could get on to the pressing business of figuring out the different occasions on which discriminating was wrong, and how to remedy these.

The US debate on affirmative action gave rise to an excursus about so-called reverse discrimination, which as far as I can tell made nobody much wiser and everybody much angrier. And this despite the fact that James Nickel early established the essential point that discriminating against blacks on the basis of their race and discriminating for blacks on the basis of their history of being discriminated against are two different forms of discrimination, and that condemning the first while condoning the second does therefore not necessarily involve hypocricy. Oddly and unfortunately this trivial observation has, as far as I can tell, not had the impact it deserved outside of the philosophical literature, or even sometimes in it. (cf. Nickel, 1972; Nagel, 1973; Thalberg, 1973; Woodruff, 1976; Simon, 1978; Singer, 1978) A second debate focused on discrimination in the labour market, where affirmative action reappared, but the possibility of indirect discrimination in particular was (and is) a contentious issue. (Levin, 1981; Wertheimer, 1983; Sunstein, 1991; Miller, 1992; Haslett, 2002) Finally, a third debate focused on so-called racial profiling, with a first round in the early '90s sparked by the publication of a highly controversial article by Michael Levin defending racial discrimination based on statistical differences, (Levin, 1992; Thomas, 1992; Adler, 1993; Corlett, 1993; Cox, 1993; Pojman, 1993; Applbaum, 1996; McGary, 1996; Wasserman, 1996) and a second, more sophisticated round, by the publication of a qualified defence of racial profiling by Mathias Risse and Richard Zeckhauser (Risse and Zeckhauser, 2004; Lever, 2005; Lippert-Rasmussen, 2006b; Lever, 2007; Levin, 2007; cf. also Gross and Livingston, 2002)

Recent work on the concept of discrimination

There is, as illustrated above, an extensive philosophical literature on discrimination, but also curious blind spots. Rarely, if ever, did the participants in these debates pause to ask that most philosophical of questions: "what does it *mean* to discriminate against someone?" In fact, it was not, I think, until the classic article by Larry Alexander in 1992 that serious philosophical work on the concept of discrimination was initiated. (Alexander, 1992) Alexander develops what he claims is a taxonomy of discrimination, although it might be more accurate to say that he develops a taxonomy of the *grounds* for discrimination. Thus, he distinguishes between discrimination on the basis of direct preferences for or against goods or people, the similar forms of proxy discrimination, which is to say indirect preferences for or against the same things, as well as several subspecies within these different types.

After this, there was an extended hiatus until discussion of the conceptualisation picked up again in recent years with the work of Lena Halldenius, Bert Heinrichs and Kasper Lippert-Rasmussen. (Halldenius, 2005; Heinrichs, 2007; Lippert-Rasmussen, 2006a; although cf. also Edmonds, 2006) Admittedly, the intervening years saw the publication of a number of encyclopaedia articles, but none of these present very satisfactory developments in the conceptualisation. (Nickel, 1998; Wasserman, 1998; Ezorsky, 2001) Lena Halldenius, however, makes important clarifications in her "Dissecting Discrimination", where she presents: "...*the Standard View*: Discrimination is decisionmaking representing (or resulting

in) a disadvantage for someone (P) on grounds that are irrelevant in the decisionmaking context (C). The ground (X) is a personal characteristic." She then convincingly argues that this conceptualisation of discrimination fails three of the six challenges she poses for it, specifically that it cannot "d. offer a stable criterion for fairness, e. be resilient against unfair background factors [or] f. contain a nonarbitrary and nonquestion-begging principle for ground selection." (Halldenius, 2005, p.457) Similarly, Bert Heinrichs helpfully summarizes a number of the key distinctions and suggests that we might conceive discrimination as occurring when "[someone] treats [person A] differently than another person B although he should not do so because they are not different". (Heinrichs, 2007, p.102) Heinrichs proceeds to argue along similar lines to Halldenius, that specifying the relevant (or irrelevant) form of differences is extraordinarily difficult. I explore all of the conceptualisations above in "But Some Groups Are More Equal Than Others".

It is worth noticing that all of the conceptualisations I have touched upon so far have considered the question of when discrimination is morally wrong to be primary, or even held wrongness to be an integral part of the concept of discrimination. Possibly it was the decision to separate the two questions of how to conceptualise discrimination and how to explain its wrongness that allowed Kasper Lippert-Rasmussen to take several important clarificatory steps in his work on the topic. Since my own understanding is strongly influenced by and largely follows his (Lippert-Rasmussen, 2006a; 2007a; 2007c), rather than introducing

his work separately in the following I simply flesh out the descriptive concept as I understand it, referencing Lippert-Rasmussen and noting any points of disagreement where appropriate.

The basics - differential treatment

Discrimination is fundamentally concerned with differential treatment. That is, discrimination pertains to actions of an agent, and is a description of a particular kind of action, which affects two groups, and affects them in different ways. Thus we may initially say that an agent discriminates *iff*

> 1) She treats those with a particular trait (T-persons) differently than those without the trait (¬T-persons)

Even this relatively simple first condition requires several clarifications.

First, the **groups** involved are defined by their traits, and 'trait' is meant to include all properties of a moral patient that could form the basis of identification of that patient as belonging to one group or another. Thus, even the arbitrary division of two groups of people into two groups ("All of *you* go over there, all of *you* go over there") assigns to them the traits of belonging respectively to the first and the second of these arbitrarily constituted groups, which traits can then form the basis of discrimination. Furthermore, the fact that one group is defined as a negation of another should not be taken to exclude situations where there are simultaneous cases

of discrimination involving more than these two groups; such situations can still constitute discrimination. Thus, a country which treats Christians better than Muslims might superficially seem to not involve discrimination on the above condition, because the group of Muslims is not identical to the group of non-Christians (we assume that there are also atheists, Jews, Hindus, etc.) and vice versa. But a better analysis of the situation is presumably to say that there are two types of discrimination occurring simultaneously – one for Christians and one against Muslims. (See below, cf. also Edmonds, 2006, , chap. 1 & 2) Many theorizers of discrimination have wanted to restrict the range of groups in some way, such as to include only "socially salient groups". I find this condition unnecessary, but explore it and argue at length against including it in "But Some Groups Are More Equal Than Others", and so will not pursue the argument here.

Secondly, the notion of **treatment** is more complicated than may appear. I explore this concept at greater length in several places, most notably in "Stealing Bread and Sleeping beneath Bridges" and "Blind Justice and a Jury of Your Peers", but for now we can note merely that it is not sufficient that the agent performs two different actions. It is also necessary, and ultimately more important, that the agent performs two actions which differentially affect the two groups. Using my left index finger to point one group of lost passers-by in the right direction, and my right to point another group in the right direction presumably does not qualify as differential treatment. But using the same index finger to

point first a sighted and then a blind person in the right direction might.

Finally, note that the apparent direction of the difference in treatment is **reversible**, that is, it is an equally valid and synonymous description of any situation meeting this condition to say that the agent treats ¬T-persons differently than she does T-persons. The condition above is simply the easiest way of formulating the fact that the agent treats one group in one way, and the other group in another, different way.

The causal link – standard and proxy discrimination

Secondly, for standard cases of discrimination we want the connection between possessing trait T and being differentially treated to be non-arbitrary for it to count properly as discrimination *of* T-persons. If a lottery happens to award prizes only to men, we do not want to say that this is discrimination of women, because the genders had no bearing on the differential treatment. Initially, it may be tempting to link the two via the intentions of the discriminator, so that discrimination is only discrimination of a group if the discriminator has intended to treat that group differently. But I think this is actually neither necessary nor sufficient. It is insufficient, because we might at the very least require that it be part of the discriminators intentions that she intends to discriminate the group of T-persons *because* they are T-persons, that is, that their possessing trait T is part of her intention, or what leads her to form the intention of discriminating. But once we recognize the necessity of this causal

link, I think it is a plausible next step to allow that even unintentional discrimination can be directed in the appropriate way so as to constitute discrimination *of* T-persons, as long as the possession of trait T plays a suitable causal role in the discrimination occurring. Thus, I recall seeing a video some years ago demonstrating that the face-tracking software of a prominent IT-company was incapable of recognizing black faces, whereas it responded well to white faces. Presumably, not only is the computer running the software incapable of forming intentions of any kind, but no discriminatory intentions need have been part of producing the software. It may have been a simple programming error, or at most the result of an unconscious tendency to design products with a specific target demographic in mind, but we might nonetheless want to say that the software discriminates black persons, by treating them differently than white persons. And it does so, because their being black plays a direct causal role in the software's treating them differently.

Thus, with some qualifications I want to adopt Lippert-Rasmussen's formulation (cf. Lippert-Rasmussen, 2007a, p.57-60; Lippert-Rasmussen, 2007c, p.823-824) of a requisite 'suitable explanation', and say that an agent engages in standard (or 'direct') discrimination *iff*:

> 1) She treats those with a particular trait (T-persons) differently than those without the trait (¬T-persons), and

2) the differential treatment is suitably explained by the T-persons possessing trait T

On the other hand, there are limits to what type of causal link will do. Although I will leave these more vague than I should prefer, at least one limitation is worth briefly discussing. Frequently, standard discrimination is contrasted with proxy (or 'indirect') discrimination. This occurs, roughly, when a group defined by possessing trait T is differentially treated, and the differential treatment is suitably explained by their possessing a different trait, T_1. Thus, women might be said to be discriminated on the labour market, e.g. by being given lower priority in competition for jobs, because they are capable of giving birth to children, taking maternity leave and are typically the primary caretakers in the family. None of these traits (with the arguable exception of the ability to give birth) are perfectly co-extensive with 'womanhood'. A company that gives lower priority to women might truthfully claim that the traits which do all of the causal work in explaining their discrimination are those above, which are instrumentally important to the company's objective of maximising the profit their individual employees generate. And they might further claim that they do not, therefore, discriminate women. Similar problems attach notoriously to e.g. educational institutions giving priority to the best applicants, which will in some contexts leave students of a particular race or ethnicity with very low chances of obtaining admission. Or, in an example Elizabeth Anderson shared with me, to the automobile manufacturer that designed the airbags of its

cars to specifications based on the average US male, with the result that the airbags severely injured people of lighter build, who happened to be predominantly women. But note that the way possessing T_1 explains the differential treatment differs. In the first case, being a woman is a proxy for possessing T_1, that is, the company differentially treats women because being a woman is likely to entail possessing T_1 (Childcare responsibilities, which is the ultimate target of the treatment). In the third case, women are disproportionately affected by differential treatment of persons possessing T_1 (being of light build), because possessing T_1 is likely to entail being a woman.

I argue in "Stealing Bread and Sleeping Beneath Bridges" that these distinctions are much less interesting than has typically been assumed. But I shall not pre-empt the argument to be developed there. Let me say merely that I roughly believe that we might want to take seriously the idea that what the above-mentioned company engages in is not discrimination against women, but that we should also seriously consider the possibility that its being discrimination against something else makes no difference to the wrongness of the discrimination. I sketch what I think is probably a better way to assess the wrongness of differential treatment based on statistical differences, in the specific context of racial profiling, in "The Art of the Unseen".

The effect - discrimination for, against and between

A final distinction bears elaboration. I argued above that discrimination is inherently tied through the concept of treatment

17

to how it affects the discriminatee(s). I have applied until now, however, the slightly clumsy concept of 'discrimination *of*', and used somewhat cumbersome formulations to allow my using it. The way discrimination affects the discriminatee(s) allows a somewhat more nuanced and easier way of talking about it, if we distinguish between those being advantaged, disadvantaged or neither by the act of discrimination. Thus, we can qualify an act of discrimination by saying that it is an act of discrimination *against* the group that is disadvantaged by it (if either group is), that it is *for* the group that is advantaged by it (if either is), and that it is *between* the two groups if neither is advantaged or disadvantaged. The last of these may seem to contradict the claim I made above that simply performing two different actions was insufficient for differential treatment. But the difference at stake here is that between performing two actions that identically affect the two groups (which is not differential treatment) and performing two actions (or one) that differently affect the two groups, though not in ways that advantage or disadvantage either. Possibly, some gender divisions are of this character, such as gender divided toilets, which it seems plausible to say differentially affect men and women, by restricting their access to two different sets of facilities, but leaves neither party better or worse off.

Summing up, I want to say that an agent engages in discrimination *against* T-persons *iff*:

> 1) She treats those with a particular trait (T-persons) differently than those without the trait (¬T-persons),

2) the differential treatment is suitably explained by the T-persons possessing trait T, and

3) the treatment is disadvantageous to T-persons

Similarly, an agent engages in discrimination *for* T-persons *iff*:

3) the treatment is advantageous to T-persons

And, finally, an agent engages in discrimination *between* T-persons and ¬T-persons *iff*:

3) the treatment is neither advantageous or disadvantageous to T-persons or ¬T-persons

The wrongness of discrimination

Having explored the descriptive sense of discrimination, we can turn now to the differing attempts at explaining what is morally wrong with it. At least four accounts of the wrongness of discrimination can be distinguished in the literature.

First, it is commonly assumed that discrimination is wrong when and if it involves treating people differently, although there is no good reason, or arbitrarily. Thus, e.g. Joel Feinberg, though he admits that part of the wrongness is explained by the harm discrimination causes (see below), argues that: "The more important part of the explanation why discrimination as such is

unjust, however, consists in its absolute groundlessness, or grounding on morally irrelevant criteria, and the characteristic sort of offensiveness these features engender, for the general characteristic this form of injustice shares with all the others is that, quite apart from any other harm, or hurt, or wrong it might bring to the one who suffers it, it offends against impersonal reason itself." (Feinberg, 1974, p.318-319) Call this the irrelevance-account, since it claims that it is morally wrong for an agent to differentially treat people when the only differences between them are morally irrelevant. And note that this is not a mental-state theory, the corresponding version of which would hold that it is morally wrong for an agent to treat people differently when the differential treatment is caused by the agent's taking into account in her deliberations differences that are morally irrelevant. I call this negative disrespect, and deal with it in "Discrimination and Disrespect" (see also disrespect just below). It is rather the principle at stake in notions that "like must be treated alike", which I explore, although in the limited context of criminal justice, in "Blind Justice and A Jury of Your Peers".

Secondly, some hold that what is wrong with discrimination (when something is) is that the treatment is unfair to the discriminatee. Call this the unfairness-account of the wrongness of discrimination. While a rather obvious possibility, to my knowledge it is only very recently that this account has been taken up by Shlomi Segall, who argues that discrimination is bad when it undermines equality of opportunity. (Segall, Forthcoming) Note that unlike the irrelevance-account, this account is concerned with

either non-comparative procedural justice (equality of opportunity) or telic comparative justice (equality of the distribuendum). I do not explore this account in the thesis. Primarily, this is due to the fact that Segall's work only came to my attention just prior to his publishing it, in the summer of 2011.

Thirdly, perhaps the most widely discussed account holds that the wrongness of discrimination is due to the disrespect towards the discriminatee that it involves. (Arneson, 2006; Hellman, 2008; Alexander, 1992; Glasgow, 2009) The background of such explanations is Kantian, and the account assumes a broad deontological principle against disrespecting the moral worth of persons. Although several variations exist, disrespect in this type of account is broadly speaking held to be concerned with the misestimation or misrepresentation of the moral worth of the discriminatee. Thus, treating blacks or women in a disadvantageous way is held to be wrong because it either presupposes or signals that these persons are moral inferiors. I explore this account at length, and argue that it is not convincing, in "Discrimination and Disrespect".

The fourth type of account, and the one that I have accepted, explains the wrongness of discrimination by the harm that it causes. Several of the accounts listed above recognize that one reason why discrimination can be wrong is that it may cause harm to discriminate, but argue that such harms are extrinsic to discrimination, that is, they are harms caused while discriminating rather than harms caused by discriminating. (Alexander, 1992; Arneson, 2006; Hellman, 2008) Full-fledged harm-accounts will

21

insist that there are harms caused by discriminating, whether they be reinforcing the social castes of a society (Edmonds, 2006), diminishing the deliberative freedoms of agents (Moreau, 2010), causing stigmatization (Wasserman, 1998) or depriving persons of goods that they ought to enjoy (Singer, 1978; Lippert-Rasmussen, 2006a).

Note that, though this is commonly assumed, there is no principled reason why there could not be a pluralistic account of discrimination. Segall argues that if there were, there would have to be a hierarchy between them, that would allow us to decide conflicts between the different wrongs, and that if there were such a hierarchy, then we could focus on whatever we established to be the most important wrong. This seems to me problematic on both counts. (Segall, Forthcoming, p.11-12) First, even if there were a hierarchy, the only situation in which we could limit our attention to the most important wrong would be if it both always instantiated in cases of wrongful discrimination *and* always trumped all other wrongs combined. And more importantly, although it would certainly be convenient for us if a plurality of wrongs happened to be arranged hierarchically, our deliberative convenience is hardly an argument that this must be the case. We might as well be unfortunate enough to be stuck with a plurality of incompatible wrongs, leaving cases of conflict fundamentally unresolvable. I am not suggesting that I think this is the case. But I believe that we would need to rule out the possibility based on the specific constitution of the wrongs involved, rather than *prima facie*.

Finally, it is possible of course to be sceptical that there is in fact *any* distinctive wrong associated with discrimination. (Cohen, 1994; Radcliffe Richards, 2000; Halldenius, 2005; Heinrichs, 2007) Such proponents will typically argue that anything which is wrong in cases of discrimination have to do with extrinsic factors, such as the fact that the differential treatment involves independent wrongs or violates an explicit policy of equal treatment and is therefore hypocritical.

What does it mean for it to be the "wrongness" of discrimination?

A central issue in exploring the wrongness of discrimination is what exactly it means for discrimination to be wrong. That is, what does it mean for the particular action of discrimination to possess the type of property that wrongness is? This is a separate and different question than either what constitutes the wrongness – the answer to which is provided by various accounts of the wrongness of discrimination – or the metaethical question of what it means for an action to be wrong more generally.

Consider the comparison with lying. Presumably, lying is a type of action of which we could give a strictly descriptive account, along the lines of "an agent lies *iff* she provides information that she believes to be false to another person with the intention of causing that person to form a false belief related to the information". However, lying is also a concept with strongly negative moral connotations, so that we might reasonably ask the question "what is the wrongness of lying?" Doing so need not mean anything

23

more than "under what circumstances is lying wrong", and could still allow for a variety of answers ranging from Kant's infamous reply – always and under all circumstances, even when telling or simply withholding the truth would e.g. aid a murderer in killing an innocent (Kant, 1797; 1993 [1797]; cf. also Kant, 1996 [1797], p.182-184) – to the strictly consequentialist that it is only wrong when doing so has worse consequences than an alternative course of action. But on any account we would need to separate the wrongness of lying from any wrongness pertaining to other features of an action, even were these features coincide with (but are not part of) acts of lying. Imagine the following scenario:

> **Soap-opera secrets**: Maria tells Ruth the sad fact that Ruth's husband Ted is deceiving her with Bethany. She also lies to her that Ted's infidelity is caused by his no longer finding Ruth physically attractive after her recent pregnancy. Suppose, plausibly, that hearing this is emotionally painful for Ruth. But suppose further that it is actually true that Ted no longer finds Ruth attractive, but not true that Ted's infidelity has any relation to his fading attraction. Rather he had fallen in love with Bethany long before the pregnancy, and his finally giving in to his feelings just happened to coincide with it.

The question is, supposing that causing emotional harm is wrong, should the wrongness of this harm be attributed to the action of lying? It seems to me that at least partially – though perhaps not

entirely – the harm is caused not by lying, though in the process of telling a lie, but rather by the revelation of an unpleasant true fact. We can call the wrongness which properly pertains to the action at stake – lying in the example just above, and discrimination in our more general discussion – wrongness which is intrinsic to the action, and any wrongness which although it occurs in the context of performing the action at stake does not properly accrue to it wrongness which is extrinsic to it.

Secondly, it is worth noting that when we are discussing what it means for discrimination to be wrong we are concerned with a pro tanto property, not an all things considered property. Thus, it is quite possible for discrimination to be wrong but nonetheless permissible, because the pro tanto wrong is outweighed by other factors. Suppose that we accept both a harm-based account of the wrongness of discrimination and a more general consequentialist theory of the moral status of actions. If an act of discrimination causes harm to the discriminatee, but simultaneously prevents much larger amounts of harm (to others, or even to the discriminatee), then it seems perfectly reasonable to say that the act is in a limited sense wrong for causing harm, but all things considered permissible (or required). Something similar will apply on deontological accounts, if we imagine a person who has promised to perform an action that turns out to constitute wrongful discrimination, and that the duty to fulfil that promise happens to weigh heavier than the duty to not invidiously discriminate. Some authors have adopted the terminology of discussing when discrimination is bad (pro tanto) versus when it is

wrong (all things considered). (cf. Lippert-Rasmussen, 2006a; Segall, Forthcoming) I have opted to follow what I think is the mainstream in the literature and use wrongness unqualified to mean pro tanto. Since I hardly touch upon all things considered wrongness, which would after all be relative to the specific and total conditions of each individual case, I do not believe that this should cause the reader any problems.

The harms of discrimination

The pressing question after considering the above, and given that I accept a harm-based account of the wrongness of discrimination, is whether there are harms intrinsic to discrimination? After all, and as seems plausible, much of the harm in those cases of discrimination that attract our attention may be extrinsic to it. Thus, e.g. depriving a particular group of the access to particular goods, such as education, because of their race or gender, involves harming these persons; it "makes the discriminatee worse off". (Lippert-Rasmussen, 2006a, p.174) Note that I concur with Lippert-Rasmussen that the relevant baseline is moralized, i.e. that an action only harms someone if the person is made worse off than she *ought* to be. But it is hardly the discrimination which is responsible for this harm, as is readily apparent from the fact that the harm to them would be largely unchanged should we decide to suddenly deprive everybody else of the same goods. If e.g. an education is a good, even if perhaps only an instrumental good and perhaps only in most – not all – cases, then being deprived of it is a harm no matter what may or may not happen to other persons. And if this is the case, then the harm is extrinsic to the act of

discrimination. (pace Lippert-Rasmussen, 2006a, p.174-178; Lippert-Rasmussen, 2007c)

But if we consider again the example, it seems that there might be *some* difference to the harm done to the deprived group if we suddenly deprived everyone else. I believe that the difference can be spelled out under several headings:

Positional goods deprivation. First, part of the benefit of education is its usefulness in accessing other goods in competition with other persons; it is a good which is partly, by virtue of this usefulness, both instrumental and positional. And though we may question whether an egalitarian distribution of intrinsic goods is inherently valuable, there seems little question that an egalitarian distribution of non-intrinsic goods is conducive of increasing intrinsic goods, or that eradicating the systematic deprivation of a positional good for one group, will promote an egalitarian distribution of non-intrinsic goods. (Wilkinson and Pickett, 2010) In less convoluted terms: if we were suddenly to end access to education for everybody, this would undoubtedly have a great many negative effects, but among its positive effects might be that one particular group no longer enjoyed systematically poorer prospects, a fact which would likely lead to a more equitable distribution of goods in other respects, a fact again which would likely lead to increased overall well-being. And the other way around, by systematically depriving certain groups of positional goods, discrimination may promote inegalitarian distribution of other goods, to the detriment of overall welfare. That is, discrimination can involve the creation or distribution of positional

goods, the creation or distribution of which can be instrumental to wellbeing, and thus subject to moral concerns of harming and benefiting.

Emotional damages. Secondly, it seems obvious that there are cases where discrimination can cause serious distress, psychological pain or damage the discriminatee's self-esteem. Such emotional damage is more likely, though hardly only possible, when the discriminatee herself has or is a member of a group that has a history of being discriminated against. And on top of the immediate harms it seems plausible that such damage can have further detrimental effects, such as the discriminatee responding by internalizing negative stereotypes, giving up her ambitions or forming anti-social attitudes.

Stigmatization. Thirdly, discrimination may cause or support the formation and spread of negative attitudes or prejudiced beliefs about the discriminatees in others. And although possessing a negative attitude or a prejudiced belief is not intrinsically bad, the spread of these is likely to increase incidents of other harms, both discriminatory and non-discriminatory.

Severed social bonds. Fourthly, whether as a consequence or independently of the type of effects covered above it seems plausible that discrimination can lead to segregation between the groups treated differently. I do not think that segregation is in and of itself bad, particularly if the groups are large enough that whatever opportunities for interaction a person loses with members of the other group can be adequately covered by

interacting instead with members of her own. However it seems possible that segregation can cause problems, making cooperation and trust more difficult. This is particularly true for persons who are members of a shared community, e.g. because they are citizens of a country and share social and political space in that country.

The above are sketches more than anything else, but I believe they are sufficient to illustrate that there are potential harms caused by the act of differential treatment, rather than by the individual treatment either of the involved groups receives.

Discrimination in the law and philosophy

Let me conclude this section on discrimination by a brief comment on the role of discrimination in philosophy and legal theory respectively. Much of the discussion of discrimination is explicitly, and often also implicitly, indebted to anti-discrimination law and the legal and political debates surrounding it. Discrimination law is itself a large and complicated field, particularly since there exist substantially different legislations and legal systems both among different countries – even among West European and Anglo-Saxon countries – and at the international level (EU, UN). (Connolly, 2006) Partly due to its complexity and partly due to my lack of training as a legal thinker I have largely stayed clear of discussing either what the pertinent law in fact is, or even the perhaps more approachable question of what the relevant law ought to be. While I have discussed discrimination in the context of the criminal justice system in "Blind Justice and a Jury of Your Peers", I do not there engage with discrimination law, but rather

with the normative principles of adjudication. However, I do touch upon the importance of the findings of philosophy for the issue of discrimination law in "But Some Groups Are More Equal Than Others". There, however, my conclusions are mainly conservative. The fact that there is no direct link between what is morally wrong and what we ought to criminalize seems fairly obvious today, to the extent that H.L.A. Hart's well-argued entreaties in "Law, Liberty, and Morality" (Hart, 1963) has, for me, the distinct feeling of running the battering ram against an open door. Thus, while I think that philosophy can do the law a valuable service by helping us all better understand the phenomenon of discrimination, in all of its variations, I do not think that this translates directly into recommendations for legislation.

Much of the philosophical work on discrimination, however, has been done by thinkers originating in the legal tradition, and whatever the qualities of their legal scholarship it often seems that the interdisciplinary venture into the conceptualisation and ethics of discrimination is a somewhat uneasy undertaking. (Brest, 1976; Wasserstrom, 1977; McCrudden, 1982; 1985; Lawrence III, 1987; Kelman, 1991; Morris, 1995; Rutherglen, 1995; Kennedy, 1997; Harris, 2002; Bagenstos, 2003; Hellman, 2008; Moreau, 2010; Altman, 2011) There are noteworthy exceptions to this phenomenon of course, including Larry Alexander, John Gardner and Fred Schauer, while Oran Doyle, Jeremy Waldron and Tom Campbell deserve at least honourable mention. (Alexander, 1992; Gardner, 1989; 1996; Schauer, 2003; Doyle, 2007; Waldron, 1985; Campbell, 1991) Unfortunately, but perhaps inevitably given this

background, the tendency has often been to ask – if sometimes without perhaps being fully aware of it – what importance discrimination law should have for how we theorize the concept and morality of discrimination. The obvious answer there is, of course, "none", given that this would seem to involve a blatant is-ought fallacy. But careful thinkers will be able to draw on the thinking in legal debates on discrimination law without falling victim to the fallacy. However, there are two more features of discrimination law that complicate drawing on the ideas developed there, which it seems to me that few thinkers have been aware of. The first is that discrimination law is, for reasons that have to do with the fundamentally liberal nature of positive law, necessarily concerned only with discrimination in certain contexts. And secondly, perhaps for the same reasons, it is only concerned with discrimination against particular groups. The second of these two limitations is the more potentially seductive and misleading of the two, and has been translated directly into conceptualisations of discrimination by at least some thinkers. I argue that this constitutes a mistake in "But Some Groups Are More Equal Than Others". But the first should not be overlooked either. As I mentioned above, there are good reasons why we do not want to criminalize all actions that are morally wrong. And I think we can reasonably expect the limitations of discrimination law to be at least partially based on and coherent with these reasons. Thus, it may be true that dinner party etiquette can involve morally wrong cases of discrimination – someone who regularly invites her neighbours for dinner, but excludes neighbours belonging to a particular group, say a racial minority, might well be doing

something morally wrong – but criminalizing such behaviour might impose more costs than it produces benefts, in the shape e.g. of unintended consequences such as damaging the social good of even non-discriminatory dinner-parties, because these are now held under the threat of legal sanctions.

What this means is that we should at the very least be careful in drawing upon discrimination law when doing philosophical work on discrimination.

The thesis and how it all fits and does not fit together

Kasper Lippert-Rasmussen once suggested to me that a philosopher need only be consistent within the bounds of any one text: no obligations to not contradict yourself in the next article. Although there is something eminently sensible about this "New Criticism"-approach to the author, I am not sure how serious he intended the statement to be taken. Nor do I expect the reader to accept it as an excuse, upon discovering glaring contradictions between two or more of the articles to follow. But hopefully, no such contradictions will emerge, and the few that might superficially resemble such will be revealed to mask perfect harmony.

The five articles that constitute the bulk of this thesis cluster round a central theme, rather than follow a linear narrative. The order in which they appear is thus somewhat arbitrary. Nonetheless, I feel that there is some benefit to the reader in arranging them as I have done. In this arrangement, they move roughly from the strictly

conceptual to the discussion of moral wrongfulness and finally to the application of moral theory to two issue within the context of the criminal justice system. This order also represents something like the order in which I have actually written the articles, with the exception that "Stealing Bread and Sleeping Beneath Bridges" was one of the last articles I began, and the caveat that there have been large periods of overlap, during which I have worked at two or three articles simultaneously, or set an article aside to work on something else, and then returned to it. The progression here will therefore to some extent, but only to some extent, mirror my own progression through working with and understanding the issues surrounding discrimination.

In the first article, **But Some Groups Are More Equal Than Others – A Critical Review of the Group-Criterion in the Concept of Discrimination**, I critically examine a standard feature in conceptions of discrimination which I dub the group criterion, specifically the idea that there is a limited and definable group of traits which can form the basis of discrimination and by implication the notion that other groups cannot. I illustrate the prevalence of this assumption and then review two types of argument for the criterion. One focuses on inherently relevant groups and relies ultimately on luck-egalitarian principles, the other focuses on contextually relevant groups and relies ultimately on the badness of outcomes. I demonstrate that the first type has problems fitting standard intuitions about groups that are central to discrimination and connecting a stable list of prohibited traits with luck-egalitarianism. I further demonstrate that the second

type can only succeed by introducing a threshold that is difficult to justify. I conclude that as neither type of argument is convincing the distinction introduced by the criterion is morally arbitrary, and as such the criterion is untenable. Finally, I suggest some of the both conceptual and practical implications of abandoning the criterion, including how it affects the wrongness of discrimination, the concept of indirect discrimination and the legal prohibition against discrimination.

In **Stealing Bread and Sleeping Beneath Bridges – Disparate Impact, Indirect and Negative Discrimination** I analyze the concept of indirect discrimination and argue that it is not merely difficult to provide a satisfactory account of it, but that once we establish one it is not particularly useful. By illustrating the difficulties the standard conception has in explaining the intuitive difference between paradigmatic cases of direct and indirect discrimination I suggest that a different conceptual difference may be at stake. I then argue first that this difference likely concerns the harm done to the discriminatees, and second that a distinction between two types of differential treatment, which I dub positive and negative, can account for the intuitive difference. In conclusion, I argue that this provides both a clearer understanding of the moral properties of cases of discrimination as well as providing indirect support for a harm-based account of the wrongness of discrimination and suggesting that our moral obligations qua non-discrimination may be more extensive than is frequently assumed.

The first two articles, particularly the second, provide indirect support for a harm-based account of discrimination. In **Discrimination and Disrespect** I examine the main rival. The disrespect-account of morally wrongful discrimination holds that the wrongness is explained by disrespect towards the discriminatee. I explore how strong the disrespect-account is by first introducing the concept of respect and distinguishing several importantly different features of it, drawing on Stephen Darwall's classic definition. Next, I briefly present Larry Alexander's bias variation of the disrespect-account of wrongful discrimination and sketch three basic challenges that the disrespect-account faces, concerning the narrow scope of the account, its counter-intuitive implications and its difficulty in specifying the relevant type of misestimation. Jointly I conclude that these make the bias variation implausible. I then review four variations of the disrespect-account, examining how each changes the argument for what is wrong with wrongful discrimination and evaluating whether they do better than the bias variation. I start with the opacity variation based on recent work by Ian Carter, which adds a constraint of negative respect on the agent, and the valuing variation based on recent work by Joshua Glasgow, which adds a requirement of appraisal respect. I argue that neither avoids the brunt of the basic challenges, and that both introduce new difficulties. I then present two accounts that deviate more radically from the bias variation, by Deborah Hellman and Richard Arneson respectively. The first exchanges recognition for expressive (dis)respect, whereas the second at once broadens the scope of relevant beliefs and imposes conditions on the background of the estimations of the agent. I

argue that although both of these are more successful at avoiding the basic challenges, each of them introduces serious new problems, which makes them ultimately no more persuasive. On the basis of this critical analysis I conclude that the disrespect-account cannot currently be said to satisfactorily explain the wrongness of discrimination.

In the last two articles, I finally approach the context of the criminal justice system. In **The Art of the Unseen – Three Challenges for Racial Profiling** I analyse the moral status of racial profiling from a consequentialist perspective and argue that, contrary to what proponents of racial profiling might assume, there is a prima facie case against racial profiling on consequentialist grounds. To do so I first establish general definitions of police practices and profiling, then sketch out the costs and benefits involved in racial profiling in particular and finally present three challenges. The foundation challenge suggests that the shifting of burdens onto marginalized minorities may, even when profiling itself is justified, serve to prolong unjustified police practices. The valuation challenge argues that although both costs and benefits are difficult to establish, the benefits of racial profiling are afflicted with greater uncertainty than the costs, and must be comparatively discounted. Finally, the application challenge argues that using racial profiling in practice will be complicated by both cognitive and psychological biases, which together reduce the effectiveness of profiling while still incurring its costs. Jointly, I conclude, these challenges establish a prima facie case against racial profiling, so that the real challenge consists

in helping officers practice the art of the police and not see that which it is useless that they should see.

And lastly, in **Blind Justice and a Jury of Your Peers – Rescuing Procedural Legal Egalitarianism from the Egalitarians** I argue that in the court of law, what might superficially appear as a tension between concerns of procedural equality and instrumental desirability is a false opposition. The only plausible understanding of the value of procedural equality is based on conceiving of it as a norm securing minimal standards of competency – a realist nod to the difficulties that afflict actual courts, rather than an independent ambition of any ideal court. Equality before the law is instrumental to the furthering of competency in practice – no more, but also no less. In making this case I begin by establishing the distinction between equality in and before the law, exploring how the concept of differential treatment hinges on a notion of likenesses, and specifying a principle of procedural legal egalitarianism and the conditions of procedural legal equality obtaining in individual cases. I then examine a common objection raised against principle of equality before the law, that it is conceptually vacuous, and argue that this is not the case. I proceed to argue, however, that on the basis of the analysis of the principle provided there is a strong case for the implausibility of a substantive moral principle of procedural legal egalitarianism, because the intuitions supporting it disappear once we remove factors supporting non-comparative principles. I examine, and reject, two potential counterarguments, and finally

suggest a way to partially rescue the principle by reinterpreting it as a consequentially grounded norm.

One issue may bear spelling out initially, although I have taken care to make note of it in the individual chapters as well. I alternate between using discrimination in a descriptive and a moralized sense in the articles (the former in "Discrimination and Disrespect", the latter in "But Some Groups Are More Equal Than Others" and "Stealing Bread and Sleeping Beneath Bridges"; neither in the remaining two articles). Fundamentally, this ought not to make a difference to the arguments at stake, and my reasons for doing so have been entirely pragmatic. Where I have focused exclusively on morally wrongful discrimination it has seemed to me easier to drop the qualifier and speak only of discrimination. On the other hand, where I have needed to distinguish the two I have maintained the division between a descriptive and morally neutral conception of discrimination and the moralized version which I then qualify. Hopefully, my doing so will make sense in context to the reader, and my indications of when and where I am doing which will serve to avoid confusion.

Normative background, and why it is not nearly as important as it could (should?) be

Roger Crisp asked me at our first meeting what my general normative position was. In a fit of exasperated honesty I replied "confused", to which he was kind enough to respond that this was a very sensible position. My confidence about a number of ethical views has increased slightly since then, so that today I prefer to call

myself "open-minded". But I think he was fundamentally right: confusion is a very sensible position, given the diversity of sophisticated views and profusion of strong arguments for very different positions.

Most of the work in this thesis takes no clear stand on a number of fundamental divisions within ethical and metaethical theory. Admittedly, I explore an issue from a consequentialist perspective in one article ("The Art of the Unseen"), and criticise the ability of a particular deontological approach to provide an account of something in another ("Discrimination and Disrespect"). And most of the discussion is premised on some form of moral realism, very loosely understood. If not, it would be either mistaken or meaningless, and as an authorial choice therefore rather strange, when I discuss how and when "discrimination is morally wrong". But doing so does not commit one to any one metaethical or moral point of view. The arguments generally proceed in a narrowly focused way, which amounts to saying "IF we assume A, then it looks as if B will follow but C cannot". Some of the arguments end up lending indirect support to certain positions and against others, by showing what positions are and are not capable of, but that is all. That is not to say that I am agnostic. And it will come as no surprise to the astute – or even the not-so-astute – reader to learn that my sympathies lie with act-consequentialism. In the following, I briefly summarise the normative assumptions that underlie, even if they rarely appear as necessary premises in, the articles of this thesis. As they are mostly tacit and take no active part in the arguments, I shall not endeavour to argue for

them at length, but I feel it behoves a scholarly work to at least make the author's assumptions explicit and sketch the most basic grounds for entertaining them.

Actions

Let me begin with consequentialism, which I take to be probably the normative assumption of which I am most certain. Partly, my convictions here are no doubt the result of my philosophical upbringing in an academic environment where the only real debate is between prioritarianism and impartial maximisation, and where value-pluralism is considered a strange fringe-phenomenon. Indeed, one of many surprises on encountering the academic world outside of Scandinavia was to discover that consequentialism is widely considered a view held only by the crazy or the evil, and it is thought quite likely that a value-monist consequentialist is both. Most people, particularly the british, are perhaps too polite to actually say so. "We just cringe inwardly", as a friendly Cambridge Ph.d.-student informed me. But the Uehiro Centre in Oxford, inhabited almost exclusively by Antipodeans and Swedes, remains a lonely consequentialist outpost in a stormy sea of Kantian liberals.

But partly the basis of my convictions goes back much further and runs much deeper. I recall reading Hume in a Gymnasium philosophy course, and our teacher arguing against Hume's moral doctrine on the grounds that it implied accepting the principle of "the end justifying the means". (Hume, 2007 [1751]) Whether this was an entirely accurate representation of poor old Hume is one

thing (it probably was not), but more importantly I recall inevitably asking the question of what on earth was supposed to justify the means, if not the ends? And receiving the shocked reply, that I was advocating the permissibility of killing infants so long as the circumstances were right, which even at the time did not seem very a convincing rejoinder (after all, what does the circumstances' "being right" mean, if not that they are circumstances under which doing so is permissible?).

Most common-sensical objections to consequentialism seem to come down to variations on this theme: One takes an incomplete or unbalanced set of consequences into account, typically because consequentialism is mistakenly assumed to constrain the relevant consequences in either time, space or patienthood. Thus, it is easy to show that consequentialism is counterintuitive if one constructs a shortsighted, nearsighted or biased strawman, but that is hardly an argument against consequentialism. If anything, it is an argument for impartiality. And if one is ready to suitably define the relevant set of consequences, then almost any intuition can be accounted for, and any theory "consequentialised". (Brown, 2011)

This feeling stayed with me in my years as an under- and graduate student, although perhaps with somewhat firmer theoretical underpinnings largely due to Shelly Kagan's excellent "Normative Ethics", and "The Limits of Morality" (the former, in my opinion, a vociferous consequentialist polemic thinly cloaked as a textbook, and no less admirable for it), as well of course as that holy trinity of utilitarianism, Bentham, Mill and Sidgwick. (Kagan, 1988; 2002; Bentham, 1996 [1789]; Mill, 2007 [1871]; Sidgwick, 1981 [1907])

And, as it turns out, contemporary consequentialism is capable of providing what seems to me entirely satisfactory accounts of important moral issues such as cosmopolitanism, climate change, animal rights, health care, punishment, etc. (Crisp, 1988; Glover, 1990; Walker, 1991; Hare, 1993; Singer, 1995; Unger, 1996; Singer, 2009; 2010; 2011) The shoe, therefore, is firmly on the other foot. If anything matters in what makes our actions right or wrong, then surely it is the consequences of the action. Consequentialism is a default position, and the onus is on the proponent of whatever supplementary factors are supposed to also affect the permissibility of our actions. That I have yet to encounter any argument to this effect which seemed to me persuasive does not mean it does not exist. But the fact that despite some familiarity with prominent non-consequentialist thinkers I have not met one does suggest that providing one is no mean feat (Williams, 2008 [1973]; Taurek, 1977; Kamm, 1991; Foot, 1992; Appiah, 1994; Kymlicka, 1996; Rawls, 1999 [1971]; Scanlon, 2000; Benhabib, 2002; Habermas, 2002; Nozick, 2010 [1974]; Habermas, 2004; Miller, 2005; Appiah, 2006; Benhabib, 2006; Darwall, 2006; McMahan, 2009). At least, that is, if it is not instead simply evidence, as another friendly non-consequentialist suggested, that like most consequentialists my moral compass is hopelessly and irredeemably askew.

A well-worn distinction within consequentialism holds that there is an important difference between the theory that what makes an action morally permissible is that it has consequences at least as good as any other possible action, and the theory that what makes an action morally permissible is that it follows a rule, which if

followed consistently has consequences at least as good as following any other rule consistently. Act-consequentialism is the former notion, and I have never found the alternative, rule-consequentialism, plausible for the simple and often repeated reasons that 1) it seems difficult to explain why "act so as to impartially maximise the balance of good consequences over bad consequences" could not be a rule, and 2) if it is, why it is not then necessarily preferable to any alternative rule. After all, for any other conceivable rule, there will be some situation where following the "act-consequentialism"-rule leads to the best consequences (as it always must), while following the alternative rule does not. (Smart, 2008 [1973], p.10-12; Law, 1999) While I realize that sophisticated disagreements on this issue exist, (Hooker, 2009; Parfit, 2011) I think the case for act-consequentialism sufficiently strong that exploring these has been pushed into the category of "things I really ought to get round to taking a look at one of these days".

A second important distinction concerns the difference between subjective and objective consequentialism, and the related distinction between doing wrong and being blameworthy. Objective consequentialism holds, as we noted above, that an action is permissible *iff* the consequences of the action are at least as good as any alternative action. Subjective consequentialism on the other hand, holds that an action is permissible *iff* the agent believes that the consequences of the action will be at least as good as the consequences of any alternative action; the belief at stake is occasionally qualified, to restrict it to what the agent could

'reasonably' believe or similar. As for wrongness and blameworthiness, the first concerns roughly whether the action is or is not morally permissible – although here contemporary moral theory branches out widely, and I recognize that there are strong arguments for speaking rather e.g. of whether the strength of reasons decisively favour or disfavour the action – and the second concerns whether performing the action makes the agent a suitable recipient of blame or praise. With respect to these two distinctions, I have accepted roughly what is often referred to today as the Scanlonian account: that wrongness and blameworthiness come apart, and that the former attaches to the action, whereas the second attaches to features of the agent who performs the action, such as e.g. her intentions. (cf. Scanlon, 2008, , chap. 4) This should come as no surprise, given my avowed consequentialist leanings, although these also mean that my reasons for adopting this account are hardly scanlonian. After all, it seems fairly obvious that blaming or praising are actions, and we must therefore subject them to the same standard as other actions, and assign blame or praise in the way which will lead to the best possible consequences. It is still worth bearing the distinction in mind, because it highlights how factors which will typically be of very little relevance for the one (such as the intentions of the agent) might be crucially important to the other. As above, I do not believe that my views here crucially influence the arguments I make, although some of the arguments could be said to lend indirect support to the views I hold, e.g. by showing the implausibility of a specific mental-state account of wrongness. Nor do I hold that, ultimately, blameworthiness has any intrinsic moral importance. Rather, as a

full-blooded consequentialist I hold that there are some features of agents when performing actions which, under normal circumstances, make them suitable recipients of blame because the action of blaming them will then have desirable consequences. Blaming a person who unwittingly and unintentionally does wrong is likely to cause at best no result and at worst unconstructive guilt; blaming a person who knowingly and intentionally does wrong is likely – or at least potentially – capable of inducing guilt in a way which will diminish that person's probability of doing wrong in the future.

Values

Things are more complicated when it comes to value theory and the issue of *how* consequences ought to matter. After all, thanks to Nozick and his experience machine, we can no longer take Hedonism for granted. (Nozick, 2010 [1974]) The fact that everybody (the problem of induction aside, I have yet to encounter an exception) feels strongly repulsed by the hypothetical offer of being placed in a Matrix-like virtual reality, in which they will believe themselves to be – and experience – living an incredibly pleasurable life, is often considered a knock-down argument. And yet, here I prefer to adopt much the same position as above. Hedonism seems to me a default position, given that wellbeing is perhaps the most uncontroversial element in any value-theory, and pleasure and pain seems an incontrovertible part of wellbeing. Even if, as some will want to hold, hedonism fails to provide the full picture, it seems strongly counterintuitive to hold that it plays no part, and that we should therefore be indifferent between two

sets of consequences which differ only – but very substantially – with respect to how much pain and pleasure people enjoy in the two.

Granted, there may be some forms of pleasure (and, perhaps pain, with the values reversed?) regarding which we have strong intuitions that they ought not to count in favour of a set of consequences, and perhaps even that if anything they ought to count *against* that set of consequences. Thus, many persons will think that among two sets of consequences, where the only difference is that in one a sadist derives pleasure from torturing a victim and in the other she does not (we suppose, naturally, that the pain of the victim and other such factors is the same in both sets), the set of consequences which does not include the sadist's perverse pleasure is if anything better. (Smart, 2008 [1973], p.25-26; Moore, 2000 [1903], p.257-259) My own intuitions in such cases are mixed, and I think it is likely difficult to keep out of our intuitions concerns such as e.g. the greater risk that the sadist will repeat her actions if it is pleasurable, but I think that even were we to grant that some forms of pleasure should not count, this would constitute a restriction on Hedonism, rather than a refutation. And if so, (suitably limited) Hedonism remains a plausible default position, to which proponents may argue that we should add supplementary values.

Most such attempts face a tough challenge in that potential supplements seem most plausible exactly when they produce pleasure (or alleviate pain) and least plausible when they do not, which raises the suspicion that we are confusing their actual

instrumental value for their supposed intrinsic value. For some suggested alternatives this is relatively easy to see. After all, if we imagine choosing between two worlds, one of which is hideously ugly and the other of which is extravagantly beautiful, but in neither of which there are any living creatures capable of appreciating the aesthetic qualities, then it really does seem to me that we could reasonably be completely indifferent between the two worlds. This, although hardly decisive, suggests that the pleasure we derive from beauty explains why we might think beauty is important. For other values, it is harder to produce counter-examples of quite the same strength. After all, we can hardly imagine two worlds with and without love, but with nobody to experience it. Thus it is at least possible, I think, that such values as knowledge, love or justice are in fact intrinsically valuable supplements, which would leave us with a pluralist theory of value.

If there are such supplementary goods, this might help us resolve the situation with respect to Nozick's experience machine: our intuitions might be accounted for in terms of the gains in hedonistic pleasure being outweighed by the loss of other values. Alternatively, we might argue that our intuitions about the experience machine are tracking the wrong things, or simply that we have intuitions which are understandable given the world we live in, but misleading in such hypothetical scenarios. This would be biting the bullet, and suggesting that in the highly constrained scenario described by Nozick we really should accept the offer of being placed in the machine, even if we might reasonably prefer differently in any realistic version of the scenario, e.g. because

doing so would place us entirely at the mercy of those controlling the machine, who could effortlessly make the experience abysmal rather than blissful. (cf. also Crisp, 2006, , p. 118-125)

In conclusion, while I consider Hedonism a fundamental part of the good, I remain open to the idea that other things *could* matter. As above, it is merely that I have not encountered what seems to me decisive arguments that they should.

Outcomes

What then to say about how to assess sets of consequences with different distributions of goods across persons? Off-hand it would seem that more is simply better, and I take the levelling-down objection and the mere addition problem to have conclusively shown that comparative concerns, such as egalitarianism and average maximising, have no place in our considerations. Thus, the notion that we should strive for equal distributions of the good has the highly counterintuitive implication that we could then have reason to make someone worse off, even if no-one else benefited in any way, if making that person worse off would 'level down', promoting equality. And the idea that we should strive for the highest average of goods has the highly counterintuitive implication that we would then have reason to prevent the addition of a person who would enjoy large amounts of good and subtract nothing from anybody, if that person would (ever so slightly) fall below the average and thus lower it. Both of these refutations are, of course, creditable to Parfit, (Parfit, 1984; 2002) and have attracted critics and debate, notably Larry Temkin

(Temkin, 1993; 2002). However, for the time being, I think it is at least reasonable to set them aside.

That more is simply better implies impartial maximisation. The main contender within consequentialist theory for this view is prioritarianism, which holds that the value of providing (or depriving) a person of a good depends on the amount of goods that person currently enjoys. The value-function of this relationship could theoretically be defined in any way, but the most intuitively plausible curve shape, the one that fits with the intuitions providing much of the strength of prioritarianism as a position, is concave. This means that if a person enjoys relatively few goods, adding or subtracting goods will have a large impact on the total value of the situation, whereas if the same person enjoys more goods, adding or subtracting goods will have a low impact on the total value of the situation. The value-function is similar, that is, to that describing the diminishing marginal utility of material goods, where we expect e.g. 1DKK lost or gained to make a larger difference to Joe the beggar than to Bill Gates the billionaire. But in that similarity lies also a potential problem. Much of the strength of prioritarianism comes from its apparent ability to better fit our intuitions in certain cases. Thus, if we imagine that we can distribute a certain and fixed amount of good, such as a moment's brief pleasure, to one of two people, the first of whom has had a very unpleasurable life and the second of whom has had a very pleasurable life, then surely it is preferable, all else being equal, to give the good to the poor first guy? (Parfit, 2002, , is the locus classicus of the argument) But the similarity of the two types

of value-functions raises the suspicion that the more abstract and less familiar of the two piggybacks on the more common. After all, we never actually distribute goods directly, but only the materials and circumstances which will allow people to obtain or experience them. Separating our intuitions here seems to me to be well-neigh impossible. Therefore, while I recognize the arguments in favour of weighing goods relative to the person's level of well-being, I remain partial to the idea that fundamentally more simply is better.

Metaethics

What finally to say about that most sensitive of topics, moral realism? As I mentioned above, my work proceeds in a way which is compatible with a very wide range of metaethical views, one where "realism" rules out, in fact, only a thin slice of views around error theory and hard non-cognitivism. This might still seem to some to rule out rather a lot, and I will be the first to admit that there are good arguments for both of these positions and against even this exceedingly broad form of "realism". (cf. Sinnott-Armstrong, 2006b; 2006a) Although the logical positivist style of non-cognitivism seems more quaint than threatening today, (Ayer, 2002 [1936]) any thinker with realist penchants should feel threatened by the problems raised by John Mackie's queerness-argument (Mackie, 1990 [1977], p.38-40), the diversity of moral sentiments (Tersman, 2009; although see Enoch, 2009), Moore's open question (Moore, 2000 [1903], p.66-69) or even that hoary old chestnut, the is-ought problem (Hume, 2008 [1739], p.293-306). But it is worth bearing in mind that what I assume, the mere fact that we can meaningfully talk about moral properties, and that

doing so refers to something, is compatible with views spanning from Cornell realism to Intuitionism, and perhaps even positions such as Simon Blackburn's quasi-realism or Jonathan Dancy's particularism.

Although my more specific views do not, therefore, ultimately have bearing on the arguments contained in the body of the thesis, I confess that it seems to me that any form of *solid* realism must rest ultimately on intuitions. This is bad news because there are good reasons to be sceptical of the value of intuitions in many or even all cases. (Singer, 2005; Sandberg and Juth, 2010) The fact that the carefully considered moral intuitions of responsible adults will be strongly affected by such "morally relevant" factors as the presence of used pizza boxes, as has been embarrassingly demonstrated in empirical studies, ought to send a cold shiver of dread down the spine of any intuitionist. (Schnall, Haidt et al., 2008) And yet, how could one ultimately prove something as fundamental as beneficence, that is, that we have at least *a* (non-derivative) reason to promote the good, if not through intuition? And what could possibly be strong enough as a counter-argument to make us seriously question the validity of such an intuition? (cf. Crisp, 2006, , chap. 3 & 5)

Part of me entertains the frail hope that, having finished my work on discrimination and finding myself therefore finally with the time on my hands to read "On What Matters" (Parfit, 2011), I shall find that Parfit has satisfactorily cleared up all these issues for the rest of us. On the odd chance that this turns out not to be true, I suppose the issue remains both open and pressing. Good thing,

51

then, that I am at the beginning of an academic career (or, so I hope), rather than the end.

Acknowledgements

I have included acknowledgments in the individual articles, which covers the specific feedback I got for each. Many, many people have helped this thesis come about, and some deserve credit in a more general way. I used to say that though academics might stand on the shoulders of giants, it was a long [expletive] climb up. I have come to realize since that we not so much stand on the shoulders of giants, as on a great human pyramid, each of whom were necessary to get to that point. The people I owe the deepest debt of gratitude not only helped me climb, but pushed and pulled me upwards when I got lost or tired. If I sometimes moved sideways or stood stock still, it is in spite of you.

Thanks to:

Jesper Ryberg for unfailing kindness and unflinching support. Jesper has helped me avoid countless embarrassing mistakes, shown me how to ask the right questions and go about getting the right answers, and put me in the position where I had an opportunity to do both. I consider myself exceedingly fortunate to have had a supervisor who challenged me, nurtured me and taught me by example. And whose achievements as a scholar are only occasionally intimidating while consistently inspiring. But first and foremost for introducing the younger, bachelor-student-me to a wholly new and wonderful way of doing philosophy during a

course in practical philosophy in the spring of 2002. I owe all of this to that.

My father, Michael, for introducing me to the astounding experience of thinking in wholly new and wonderful ways at all. Astral bodies, systemic therapy, hive-mind intelligence and Gaia-theory may have been challenging concepts to struggle with for an adolescent, but discussing them with you, sharing your enthusiasm, and your love and support while doing so, sparked a fire that has never gone out. I owe everything to that.

Carolina, for bearing with and caring for me through all the travails of philosophy, a burden I suspect few people realize the weight of. I do not know where I would be today without your love, but I know that the place I am is infinitely richer and happier. Hey, that reminds me of an interesting thought-experiment. Parfit has this scenario, where…

Thomas Søbirk Petersen, for innumerable morale-boosters, tons of good advice and great company. And for helping me have the courage to say "I am a hedonic act-consequentialist" with head held high. Bring it on, Rawlsians!

Jakob Holtermann for setting a sterling example of what an ambitious ph.d.-student and fantastic colleague should be. That I have yet to forgive you for going off to the faculty of Law at Copenhagen University is your own responsibility for the very same reasons. And for believing in and collaborating with me in

carrying out the crazy idea of editing the best Danish book about philosophy that nobody ever got the chance to read.

Sune Lægaard, for helping me understand the way the Kantian mind works and accept that they are not necessarily mad, even if I disagree with them. And for encouragement, helpful suggestions and good cheer.

My many other wonderful colleagues at the department of Philosophy & Science studies: Klaus Frovin Jørgensen, Stig Andur Pedersen, Erik Bendtsen, Søren Riis, Kasper Risbjerg Eskildsen and Aksel Haaning are all part of what makes the department a stimulating and enjoyable place to work. And their input on such issues as probability, free will and epistemology have frequently enlightened, or at the very least forced me to come up with better arguments for the things I have always assumed.

My once and current fellow ph.d.-students, Pelle Guldborg Hansen, Kira Vrist Rønn, Inge Schiermacher, Kasper Mosekjær, Claus Festersen, Martin Mose Bentzen, Muneza Rehman and Jens Ulrik Hansen have shared with me the big and small ups and downs of life as a ph.d.-student. Doing so has made the road less lonely, and seeing some successfully defend their theses has made the goal seem more achievable.

Kasper Lippert-Rasmussen for patience, kindness and good-humour as I blundered round the modern discrimination literature, much of which carries his signature. For numerous instances of completely supererogatory help and advice. And for being good

enough to not write any articles that happened to overtly prove me wrong while I was at it.

Nils Holtug, Morten Ebbe Juul Nielsen, Søren Flinch-Midtgaard and Claus Hansen all make the small world of Danish analytical ethics a funnier and more interesting place to be. And I am grateful to all of them for warmly welcoming me into it.

Roger Crisp, my academic advisor during my stay in Oxford, helped me find my way around the alien world of a very different and sometimes Byzantine university culture. And in both discussion and writing sets a standard of meticulous reasoning that I fear I will never achieve, but have resolved always to strive for.

Barbro Fröding, Julian Savulescu, Thomas Douglas, Lisa Makros, Deborah Sheehan, Rachel Gaminiratne, Nicholas Iles, Nancy Patel, Katharina Berndt, Alexandre Erler, Eleanor Grant, Lena Groeger, Bridget Williams, Francesca Minerva, Pablo Stafforini, Johann Frick, Ian Carroll, Anders Sandberg, Nick Boström, Anders Herlitz, Janet Radcliffe Richards and Simon Rippon (in no particular order) all contributed to making my stay at the Uehiro Centre for Practical Ethics in Oxford during the spring of 2010 one of the most wonderful and rewarding experiences of my life. I am deeply grateful for everything that I learned and the great times I had there.

David Rhys Birks for introducing my hillbilly self to many, many aspects of the modern, civilized world of international philosophy, for showing me round the Manchester Political Theory circles and

for being such an all-round excellent bloke. There will be a return to the Purple Turtle!

The coordinating group of the twin organizations NNPE/NNPT, Raino Malnes, Jakob Elster, Lene Bomann-Larsen, Mats Lundström, Eva Erman and Kasper Lippert-Rasmussen put together a series of seminars which proved to be some of the most intensive learning experiences I have had. Participation there not only gave me invaluable insights into the ongoing research of others and thorough feedback on my own work, but allowed me to meet a fantastic crowd of colleagues from all of the Scandinavian countries.

Many of my other colleagues at the university of Roskilde, in particular the often overlooked technical staff who made it possible to pursue this project without the gears of bureaucracy grinding to a screeching halt: Camilla Lomholt, Karen Kleis, Vibeke Mortensen, Vibeke Olsen, Susanne Haslund, Anya Strøm and above all Marianne Sloth Hansen. All of you deserve thanks for unfailing and friendly help with problems big and small.

My dear friend and colleague Janus Mortensen, from and with whom I learned many of the skills that made me able to both start and complete this ph.d. And whose friendship and thinking continues to challenge and encourage me across our disciplinary divide.

Last but not least, my friends and family, Else, Rasmus, Jette, Marcus, Thomas, Thomas and everyone else, for bearing with life

around a philosopher, for helping me remember that there is life around and outside philosophy and for making sure that I occasionally enjoy it. Hey, that reminds me of a thought-experiment. Kagan has this scenario, where…

But Some Groups Are More Equal Than Others – A Critical Review of the Group-Criterion in the Concept of Discrimination*

Introduction

In this article I aim to discuss what I consider an underappreciated problem in the conceptualisation of discrimination, to wit the limitation to particular groups as part of the definition.

That some form of grouping, and the divisions between people this implies, plays a necessary part in the definition of discrimination is obvious, in that the basis of discrimination is differential treatment, which presupposes distinguishing between those to be treated one way and those to be treated another. Any way of doing so may be said to rely on dividing people into groups, even if inexplicit and unreflective. Using groups in this rather trivial sense is uncontroversially necessary to the definition, because unless such distinctions are drawn no form of discrimination, even understood in its widest, non-normative sense, would be possible.

But it is not this trivial sense with which I am concerned here. My concern is rather what I shall call the "group-criterion": the idea, prominent in both legal and philosophical definitions, that *particular* groups are the subject-matter of the concept of discrimination, that these can be established prior to any specific case of discrimination, and, most importantly, that not all groups can be subject to discrimination. Typically, this condition is

expressed in the form of what we might call "the prohibited list": a selection of traits that must not be the basis of disadvantageous differentiation. Let me take as a clearly formulated example suitably close to the common-sense understanding the definition of the European-wide EU anti-discrimination initiative, where discrimination is said to occur: "…when a person is treated less favourably than another in a comparable situation *because of their racial or ethnic origin, religion or belief, disability, age or sexual orientation.*"[1] (Campaign, 2003, , my emphasis) Note that, although the definition is not phrased in the form of a strong conditional (*iff*), I think it is clear that it means to imply that discrimination does *not* occur when a person is treated less favourably than another in an otherwise comparable situation because of any trait not mentioned, so that the prohibited-list serves to exclude those traits that cannot form the basis of discrimination.[2]

Although the group-criterion has a certain common-sense appeal, it is somewhat surprising how uncritically the notion that

---

[1] The traits included here constitute something like the standard prohibited list, with the exception that gender, normally an obvious candidate for a prohibited list, is curiously absent. (Cf. Connolly, 2006, p.21-24, 39-57). See also e.g. the EU Directive on Equal Treatment (Implementing the Principle of Equal Treatment between Persons Irrespective of Racial or Ethnic Origin, 2000), the UK Equality Act (Equality Act 2006, 2006), or the European Convention on Human Rights (Convention for the Protection of Human Rights and Fundamental Freedoms, 1950), albeit such lists sometimes appear with the vague opening of including "any other grounds", subject to judicial discretion.

[2] If the strong conditional was not intended, one would expect the definition to include an "e.g.", signifying that the traits mentioned are only the most prominent candidates.

discrimination must apply to a particular set of traits, and only to those, has been adopted in the philosophical literature. Although it has received little explicit support it is frequently taken for granted, or understood as necessary even when discussed critically. Most discussions of discrimination prior to the 1980's consistently equate discrimination with racism and misogyny – religious discrimination making intermittent appearances – in line with public concerns at the time (cf. e.g. Brest, 1976; Wasserstrom, 1977). Although this focus does not strictly speaking commit such authors to limiting discrimination conceptually, the lack of qualifications seems to me to suggest as much. It is, I think, only as public debate gradually expands to include e.g. homosexuals, ethnic minorities and the disabled among the groups potentially in need of protection from discrimination that the question emerges of how many and which groups ought to be on the prohibited list.

In more recent articles the group criterion is often implicitly endorsed. Let me illustrate this with definitions from three prominent philosophical encyclopaedias: the Routledge Encyclopaedia of Ethics, the Routledge Encyclopaedia of Philosophy and the Academic Press Encyclopaedia of Applied Ethics. The first article is said to concern: "Discrimination *against* a group of persons that marks them out for unfair, harmful treatment."[3] (Ezorsky, 2001, p.412) Although this initial definition

---

[3] The definition is problematic for other reasons. On standard accounts of discrimination, discriminating against someone does not "mark them out" for unfair, harmful treatment – discrimination *is* unfair, harmful treatment. Thus,

sounds suitably broad and the scope of the concept is never explicitly defined, it is soon clear from the consistent references to gender and race that the author equates discrimination with unfair, harmful treatment of these particular groups, as when she concludes from a more abstract analysis that: "...the adverse impact characteristic of institutional discrimination may plausibly be termed *racist* or *sexist* impact." (Ezorsky, 2001, p.412) In the second we get a more explicit version of the criterion, when the anti-discrimination principle is said to require that: "When distributing educational opportunities and jobs [*list of items*], [we] never exclude whole groups of persons or choose one person over another on grounds of race, ethnicity, religion or race [*list of excluded characteristics*]..."[4] (Nickel, 1998, , italicized brackets in original) The prohibited list, Nickel recognizes, could be expanded: "...to include national origin, political beliefs, being a non-citizen, sexual orientation, income and age", but although no specific

---

the act of discrimination is not a form of designation; rather, labelling someone as belonging in group A, and not in group B, must be done prior to discriminating against them, so as to enable discrimination to take place at all. It makes more sense, I believe, to understand the marking out to relate to the group-belongingness in a causal manner, so that in effect discrimination against a group of persons consists of unfair, harmful treatment of them *because* they are marked out as members of the group in question. I return to this below.

[4] The inclusion of an item list strikes me as another strange addition to the concept of discrimination. Certainly this could make sense if we were discussing a legal principle, but Nickel is explicitly formulating a *moral* principle. The reason, as he explains it, is a liberal concern with not extending the requirements of the principle into the private sphere of e.g. selecting one's friends. But this limitation seems to require a more foundational normative principle as justification, which would make an item list superfluous, by delineating the scope of the anti-discrimination principle prior to and independently of such a list.

reasons are given for this particular selection, the list is certainly not open-ended, nor are any of the groups already on the list subject to potential removal. (Nickel, 1998) Finally, in the third: "A person is said to discriminate if she disadvantages others on the basis of their race, ethnicity, or other group membership." (Wasserman, 1998, p.805) Here, although Wasserman pursues an interesting discussion of the scope of the concept, to which I will return below, the "group membership" is non-trivial, even if, as Wasserman notes, this raises serious difficulties, in the form of: "...the obvious and controversial questions of what groups besides racial and ethnic ones can be discriminated against, and of why adverse generalizations involving those groups are particularly objectionable."[5] (Wasserman, 1998, p.807) Thus, conceptualisations of discrimination by and large remain wedded to the idea that we must limit discrimination to being, as Richard Arneson puts it: "...responsiveness of the wrong sort to *certain classifications of persons.*"[6] (Arneson, 2006, p.795, my emphasis)

---

[5] Some ambivalence is perhaps indicated by the use of "particularly", as this, on one reading, implies a much less demanding criterion, one where differential treatment against non-prohibited-list groups would be discrimination, just not particularly wrongful discrimination.

[6] Note, however, that Arneson moves from seemingly endorsing the group-criterion to abandoning it. (Arneson, 2006, p.795-796) One reason for this is the apparent difficulty of establishing the common denominator of the different groups: "However, it is tough to say what renders certain classifications problematic or especially apt for running afoul of a correct antidiscrimination norm. This becomes clear when we extend the list of classification types past supposed race and skin color. Prominent candidates include ethnicity, sex, religion, age, disability status, and sexual orientation. The common thread, if

Indeed, the articles are representative of three all too common approaches to the issue, in that the first simply assumes a given prohibited list, the second merely considers which groups to add to the list and the third goes further only to question which criterion to employ in deciding whether a group should be on the list or not. The issue that is discussed, when an issue is raised at all, is not whether there should *be* a group-criterion, but how, given the difficulties involved, to compile the list.

My question is what reasons can be advanced to support limiting discrimination to predefined groups in this manner? On the one hand, it will initially strike many people as odd to hold that someone could be the victim of idiosyncratic discrimination, such as e.g. that based on the number of syllables in their first name (equal to or less than two versus more than), or the numerical sum of the numbers in their year of birth according to CE-reckoning (odds vs. evens). On the other hand, if the reasons for applying a group-criterion will not stand critical scrutiny, the distinction is arbitrary and would appear itself to constitute a form of discrimination against all those groups excluded who thereby do not enjoy the protection granted by the norm against discrimination. After all, it is not immediately obvious why the wrong done to someone who is discriminated against, e.g. by being fired from her workplace for belonging to a certain group, should *ceteris paribus* be any less serious when that group is "those with

---

any, is not so easy to discern." (Arneson, 2006, p.795) It is not entirely clear from the text what his reasons for the initial endorsement are.

three syllables in their first name", than when it is the group blacks/muslims/gays/etc.

I shall argue that this is precisely the case: the reasons for applying a group-criterion will not stand critical scrutiny, and as such it must be dropped from the definition. Just as in the Orwellian fable I have paraphrased for my title, we should be highly suspicious when someone suggests that some groups are more equal than others.[7]

In part 1, I examine whether it is possible to justify the group-criterion by arguing that there are some groups that are *inherently relevant* to the concept of discrimination.[8] I move from intuitions about relevance to luck-egalitarianism as it relates to immutability of the trait in question, arguing that no viable argument has been offered yet for the inherent relevance of particular groups. In part 2, I discuss whether some groups might nonetheless be *contextually relevant* given socio-historic circumstances, examining the role of identity and the additional harms-argument. I dismiss these as insufficient to establish the conceptual difference required by the group-criterion, and conclude that it is unsustainable. Finally in

---

[7] George Orwell, in what has always seemed to me a stroke of authorial brilliance, has the villainous pig Napoleon supplement article 7 of the *Animal Farm*'s constitution – "All animals are equal" – with the smoothly poisonous "but some animals are more equal than others" in his 1945 novel.
[8] Note that in the present context I use relevance in the strongly constricting sense where non-relevance equals irrelevance, i.e. implies falling outside of the scope of discrimination as per the group-criterion.

part 3, I tentatively discuss some of the implications of discarding the group-criterion, both for practice and for the definition.

What's in a definition?

Before diving headlong into the arguments, let me say a few words about what I take it to be to discuss the definition of a concept like discrimination.

One important consideration is how to balance our considered views about the concept with broader or more common linguistic practices. Some might hold that we should frame our definition of discrimination in line with common usage of the term, and that since it is not ordinarily used except in connection with particular groups, the group-criterion is part of the definition. Kasper Lippert-Rasmussen's influential analysis of the concept of discrimination proceeds roughly along these lines, when he takes our common usage of discrimination, in which, he claims, we limit it to 'socially salient groups', as a starting point and proceeds to explore when and if such differential treatment is morally bad. (Lippert-Rasmussen, 2006a; 2007a; 2007c) A parallel point is made by Joshua Glasgow in a recent article on racism, where he suggests that definitions that run counter to and aim to revise common usage must count such lack of fit as a "cost" of these definitions. (Glasgow, 2009)

While I am sympathetic to these points, I believe that if framed as an objection to the enquiry that I wish to pursue in the present article it would be mistaken. Firstly, I am sceptical that common

usage is sufficiently restrictive to support the group-criterion. After all, even if it is true that 'discrimination' is most commonly used in connection with particular groups, this seems a contingent rather than a necessary feature of our usage. Few people would, I believe, bat an eyelid were I to refer to my being fired because my employer dislikes persons whose first name starts with an "F" as a case of discrimination. Not so for other parts of the definition. So that if we imagine e.g. that I was not treated differentially, disadvantaged by the treatment or no agent carried responsibility, we might well expect people to wince at the label. Secondly, presumably, if discrimination is a normatively interesting concept it is so because it points to a particular kind of wrong associated with a particular type of action, one which does not rely on how the term "discrimination" happens to be used in a particular context.[9] One need not be a Platonist to believe that those types of actions which we now commonly refer to as discrimination were morally wrong, had the same features in common and were wrong *because* of (some of) the features they share, before the term was ever applied to them. Nor to agree that whatever makes what we

---

[9] This is not, I should stress, a requirement that the concept be moralized, that is, that it is a necessary a part of the definition of discrimination that it is morally wrong. Although I believe that an argument can be made for moralizing the definition based on how the word is commonly used, and indeed that is how I use it in this article, my claim is the more modest that if discrimination is normatively interesting it must be because there are particular wrongs contingently associated with it. That is, it must be true that discrimination *can* be wrong for reasons that have to do with the particular type of action it constitutes. I return to this point briefly in the section on contextually relevant groups below.

actually happen to *call* discrimination wrong will, barring a morally relevant difference, make similar actions wrong for the exact same reasons, even if we happen to not ordinarily call such actions discrimination. Whether the group criterion is based on such a morally relevant difference *is* the issue I wish to examine. Of course, we could, as a matter of pure terminology, collectively decide to reserve the term "discrimination" for the kind of action directed at particular groups, but on top of being arbitrary this seems needlessly complicated and potentially confusing if there are related phenomena that share all relevant features.[10]

So if not a fit with common usage, what are we looking for, when we consider whether to include the group-criterion in a definition of discrimination? A different approach will take a definition suitably close to the way the concept is used in ordinary language as its starting point and then evaluate the extent to which this usage maps onto the normative contours of the phenomenon the concept is meant to encompass. My argument in the following proceeds along these lines towards the conclusion that no explanation of the wrong involved is consistent with a tight fit between normative distinctions and the element of our ordinary usage which constitutes the group-criterion. This is revisionary, but at the gain of removing a potential conceptual obfuscation, i.e. the

---

[10] If English was capable of diminutives we might then apply such to the idiosyncratic part of the concept, as in the German neologism "Diskriminatiönchen".

possibility of overlooking wrongs identical to those identified by the concept of discrimination group-criterion-included, because they happen to be not included in our ordinary usage.

Let me demonstrate by way of example how I intend to proceed. Perhaps the most commonsensical way of justifying the group-criterion is by arguing that there are some traits which are determinate of the wrongness of what is going on when someone discriminates. Discrimination is wrong, on this account, because it involves treating A differently *because* A is X (where A is a person or a group of persons, and X is a trait on the prohibited list). Thus, many of us will feel intuitively e.g. that it is profoundly unfair to treat a woman different than a man *because* she is a woman. Gender is important because it is constitutive of the wrong perpetrated in a way that having three first-name syllables is not. There are, however, two problems with this understanding, which helps us illuminate the requirements for a definition.

The first problem with the common-sense understanding, as I have portrayed it here, is that the differential treatment on which it relies is insufficient to establish the wrongness. We might agree that in some sense of differential treatment, such as e.g. disadvantage, we only discriminate when someone is differentially treated, but if a necessary this is hardly a sufficient condition. We all of us differentiate between people on a daily basis, and in myriad ways which clearly prove advantageous to some and disadvantageous to others: If I invite four friends to dinner, I have given to them the good of a free meal and (so I hope) pleasant company, whereas I have not given to others, friends and strangers

alike, those goods, and so clearly have put them at a disadvantage compared to the lucky invitees. But few of us would think that this form of disadvantageous differentiation and the innumerable examples like it are wrong, or at least not that they are *all* wrong. And it does not seem to make any difference to such cases if e.g. I invite my four male friends to dinner, but not my female friends, because I feel like a "boys' night". Although this is clearly selection on the grounds of gender, calling it discrimination, in the pejorative sense we are concerned with here, does not seem right.[11] We need to limit discrimination to those cases where differential treatment on the grounds of X is morally wrong.[12] Thus, in the following, I adopt roughly the definition that discrimination against occurs when an agent differentially treats two groups, because of one agent's possessing trait X, which the other agent does not possess, and the differential treatment is morally wrong.[13]

---

[11] Note that the requirement should be understood only to be that discrimination is *pro tanto* morally wrong, not that it is morally wrong all things considered. Thus, the claim is that there are plausible cases, such as the discrimination in favour of four friends invited for dinner, where discrimination does not even in and of itself constitute a wrong, irrespective of whether there might also be additional reasons that would have outweighed the reasons not to discriminate had it been wrong.

[12] What, exactly, it is that makes discrimination morally wrong is a contentious issue. Current debate tends to divide into disrespect- and harm-based accounts. (Radcliffe Richards, 2000; Halldenius, 2005; Arneson, 2006; Edmonds, 2006; Lippert-Rasmussen, 2006a; 2007a; 2007c; Heinrichs, 2007; Hellman, 2008; Glasgow, 2009; Moreau, 2010; Altman, 2011; Segall, Forthcoming) For present purposes, however, I need take no stance for or against a particular account.

[13] The definition I here employ is heavily indebted to, although less sophisticated than, Lippert-Rasmussen's work in (Lippert-Rasmussen, 2006a). Note, though, that in the present context I deviate from Lippert-Rasmussen's

The second problem with the common-sense understanding, at least in the form I have given it above, is that its central claim – that differential treatment of A is morally wrong if it is because A is X – is flagrantly question-begging. It does indeed appear that any argument for the group-criterion would need to rely on a moral difference of this kind, but what we require is an explanation as to *why* gender, or any other trait, should have this special status, not merely an assertion that it is so. Note however the difference between a justification of the group-criterion and a *wrong-making principle*. Any explanation of what is wrong with discrimination will need to rely on a moral principle, which as a matter of course will distinguish between those situations in which differential treatment is morally permissible and those in which it is discrimination, i.e. morally wrong. Thus a harm-based account will hold that discrimination is wrong when it harms the discriminatee, e.g. by stigmatizing her, damaging her self-esteem or depriving her of goods which she ought to have enjoyed. In a certain sense, this constitutes a form of prohibited list with one trait: liable to be harmed by the differential treatment. And again, this trait could be

---

use, common in much of the literature, of qualifiers such as "wrongful discrimination" and follow Lena Halldenius in simply reserving the term "discrimination" for the morally wrongful type of action. (Halldenius, 2005) Doing so more generally is defensible I think – my experience is that most ordinary language users are confused by the suggestion that there can be morally permissible cases of discrimination – and I do so primarily for pragmatic reasons. As morally benign forms of discrimination (permissible (dis-)advantageous differential treatment) fall outside the scope of the article, I trust that this pseudo-abbreviation shall not cause undue confusion; it should not, in any case, make any difference to the argument I pursue.

said to be defined prior to any actual cases of discrimination. My claim, therefore, is obviously not that operating with a form of group-criterion in *this* sense is untenable, but rather that there is no list which at once resembles the standard prohibited list and is derived from such a principle in a coherent fashion. In conclusion, what the weakness of the common-sense understanding reveals is that an argument for the group-criterion will have to at once explain what justifies the distinction, and be capable of drawing a reasonably clear boundary between traits on the prohibited list and traits that should not be there. We are looking for a reason to circumscribe the concept of discrimination in a particular way, because the wrongness is linked to the quality of particular traits.

What degree of resemblance to the standard prohibited list should be considered sufficient constitutes a final complication. The question is how closely the boundary should be drawn to a way which fits with the traits generally considered to be on the prohibited list (race, gender, religion, etc.) and those not to be included (such as number of first-name syllables). If it fails in this, that is if it ends up including too little, too much or entirely the wrong traits, it does lose intuitive plausibility. Additionally, the practical interest in an argument which failed to include any of the traits we normally consider important to social justice, and thus to prohibit racism, homophobia, misogyny, etc., might likely be limited. Even so I do not think we should overstate this requirement. Arguably, if we had solid grounds for why a particular set of traits should be considered inherently more relevant we might wish to base a prohibited list on these traits,

even if the set was radically different from the set commonly conceived (gender, race, religion, etc.). If the arguments supporting a version of the group-criterion were sound, there is no obvious reason why similarity to the list of traits we tend to consider relevant should be considered a *sine qua non*.

Introductory remarks aside, let me turn now to the examination of the various arguments for the group-criterion.

Inherently relevant groups

A common-sense way of justifying the group-criterion is, as I have tried to illustrate above, by arguing that there are some traits which are inherently important when it comes to discrimination and others that are not. To avoid being question-begging this relevance must be explained. Attempts to provide such an argument essentially break down to arguments about the particular irrelevance of some traits to decision-making. This view is argued for example by Harry Frankfurt, when in his critique of egalitarianism he holds that, rather than focusing on states of outcome: "Treating a person with respect means, in the sense that is pertinent here, dealing with him exclusively on the basis of those aspects of his particular character or circumstances that are actually relevant to the issue at hand."[14] (Frankfurt, 1997, p.8) However, as

---

[14] Frankfurt's argument is not explicitly concerned with discrimination, but with the related concepts of equality and equal treatment. In explicating these, however, I believe Frankfurt in effect provides at least the initial steps in an argument for the disrespect-based account of discrimination.

Nickel observes, when viewed on the face of it "relevance" as such does not seem to be a criterion that will give the right answers, at least on a broad, instrumental construal of relevance. Excluding traits on the basis of instrumental relevance is too context-sensitive to be capable of generating a stable list, much less one that resembles the standard list: "What makes [excluding on the basis of irrelevance] plausible is the fact that one's religious beliefs, for example, have little bearing on one's qualifications to work in a construction company. Choosing people for construction work on the basis of their religion just seems irrational. But if the owners of a large construction company seek to employ only fellow Christians because they think they are more likely to be honest and hard-working, or because they want to create a certain religious atmosphere within the company, they cannot be faulted for making arbitrary choices or using irrelevant criteria. The problem is rather that these selection procedures are inappropriate for a large company in a diverse country because they are unfair to non-Christians." (Nickel, 1998) If circumstances can determine whether or not a trait can legitimately be the focus of discrimination, then the prohibited list becomes something more like a rule of thumb, and the relevance or irrelevance of any trait becomes dependent not only on context but on individual preferences. (cf. also Halldenius, 2005, p.459-460; Radcliffe Richards, 2000, p.154-155)

A promising-looking answer to this problem is to adopt a more limited notion of relevance, such as *moral* relevance, which might be what Frankfurt has in mind when he qualifies relevance with "actually", and is certainly suggested by Nickel's considerations of

fairness. Explaining irrelevance in terms of a moral principle such as unfairness will undoubtedly weed out a number of problems – it will narrow the field of preferences which determine relevance. But it is not immediately apparent that shifting the focus to a moral principle can generate a stable list of traits either. This presupposes that we are capable of determining that there are some traits which when used as the distinguishing marks in differential treatment of different individuals will make that action discriminatory, e.g. by virtue of being unfair, in all circumstances.[15]

The standard argument to this effect in the context of discrimination, featuring prominently in legal theory, focuses on responsibility for possessing the trait, often formulated as a question of the immutability of the trait, as the quality of those traits which can give rise to the distinctively discriminatory wrongness. Now, while this line of argument is generally critically received in the literature it is given relatively short shrift, and may warrant a brief, closer look.[16] (Levin, 1992; Wasserman, 1998;

---

[15] Note that this is required, because only a stable list will support the group-criterion. Therefore the plausible suggestion of e.g. relying on a notion of fairness that is context-sensitive, which is to say that it determines fairness through the appropriateness of differentiating on the basis of *this* trait in *this* context, will establish a wrong-making principle rather than a prohibited list and as such will not do. For a fairness-based account of the wrongness of discrimination, see (Segall, Forthcoming).

[16] This also means that a certain amount of reconstruction and speculation is necessary, so as to conceive of how a case for immutability *might* look, with all the dangers and difficulties implied by such interpretive exercises. I believe however that the problems I identify in the following will afflict any version of such an argument, so my analysis is not heavily dependent on correctly reconstructing any specific account.

Halldenius, 2005; Heinrichs, 2007) As Lena Halldenius describes it: "The idea is that it is particularly bad to be disadvantaged because of a characteristic one cannot help having."[17] (Halldenius, 2005, p.461) The thrust of this argument relies on our notions about the relation between fairness and responsibility, specifically the argument has intuitive plausibility because it fits with widespread sentiments about the difference between being disadvantaged due to circumstances which we have choices about, and being disadvantaged due to circumstances we do not. Thus, if somebody is refused a job because she lacks various technical qualifications, one reason this will strike many people as not unfair is because most people are, at least partially, responsible for which qualifications they have.[18] My current qualifications are (partially) the result of choices I have made, and if I do not have the qualifications necessary for a job that I desire, I can attend classes and obtain them. But if I am refused a job because I am black or

---

[17] Or, as Wasserman puts it in his argument for discounting mutable traits: "However, annoying or offensive it is to be disadvantaged because of one's "membership" in an accidental or transient group (for example, to be frisked because one happens to be on a crowded subway car just after a gun is fired), this hardly counts as discrimination." (Wasserman, 1998, p.807) Note that both of these accounts allow that there can be other reasons why an action is morally wrong that apply to a case of discrimination. Whether the wrongness traceable to immutability is then conceived as a threshold-argument or an argument for an entirely separate type of wrongness is not clear.

[18] This will be less true the more difficult it is for the average person to obtain the requisite qualifications, such as the far too common situation where gaining access to the education providing these qualifications is expensive enough that it is essentially out of reach of some members of society, but I do not mean to suggest this example as illustrative of a universally valid principle, only to explore the thrust of the intuition.

female, then I am disadvantaged even though there is nothing I have done to bring the state of affairs about, or can do to alter it.

If we turn to the analysis of how this line of argument could work several things bear mentioning. First, note the crucial ambiguity well captured in Halldenius' formulation that the relevant traits are traits "we cannot help having". As Bert Heinrichs points out and the example above illustrates, this could mean both a future- and a past-oriented lack of responsibility.[19] (Heinrichs, 2007, p.104-105) This distinction is important because although the two possibilities can rely on similar moral principles, the way they point in different directions will have implications for the contents of the prohibited-list. The past-oriented interpretation relies on the idea that it is unfair to be disadvantaged (or advantaged) because of possessing a trait which one is not responsible for coming into the possession of. The alternative, future-oriented account, does not track my past responsibility for possessing a trait. It concerns rather my responsibility for possessing it now and in the future, in the sense that it assesses whether possessing it now and in the future is optional or a condition imposed upon me.

---

[19] Heinrichs is, to my knowledge, the only one to draw this useful distinction, but he fails, I feel, to fully develop the normative implications of it. This may be due to the fact that his argument proceeds in the opposite direction, from the rejection of fixed criteria due to their inability to accommodate contextual relevance, to what seems to me an essentially relativist conception of discrimination.

The immutability-argument, strictly speaking, concerns the (postulated) unfairness of being disadvantaged because one possesses a trait which one cannot alter, no matter how one came to possess it, that is future rather than past responsibility.[20] And although the two will sometimes coincide, as they do in the example for the trait "race", it is quite possible both to carry past-oriented responsibility for obtaining an immutable trait and to not be so responsible for obtaining a mutable trait. Developing an example in Heinrichs, we could say that the first is the case if I drive recklessly, crash and suffer an incurable disability, whereas the second is the case if I become disabled through an accident whose occurrence I have no responsibility for, but am then given the choice of undergoing a therapy that will remove the disability. (Heinrichs, 2007, p.105) This leaves us with three versions of the argument: past-responsibility, future-responsibility and a combination of the two. However, as I shall argue, only the immutability, and therefore the future-responsibility account, could even hypothetically support the group-criterion, even if it too is not ultimately capable of plausibly doing so, in addition to which all three suffer from common problems.

---

[20] This strict sense is not, however, the way the immutability-argument is normally presented. Rather, it is either paired with or understood to imply traditional luck-egalitarian norms. The idea in the later case seems to be that immutability, while not morally relevant in itself, will serve to pinpoint those traits which we can be certain that agents are not responsible for possessing. See e.g. Larry Alexander (Alexander, 1992), quoted below.

Let me start with those issues common to the three. One initial problem, which is often taken to be decisive, is that in none of the three versions does the argument fit particularly well with what we consider the standard prohibited list. (cf. e.g. Halldenius, 2005) The argument is by turns both insufficient and unnecessary to providing a list that resembles our intuitions.

It is insufficient because there are an unlimited number of traits that will meet any and all of the versions, but are not regularly considered to be constitutive of groups that can be subject to discrimination. The numerical sum of the numbers in CE-reckoning birthdate that I used as an example earlier would be one. As such, any version of the argument will either require additional conditions that can narrow down the field, if it is to approximate the standard prohibited list, or accept expanding it essentially without limits.

It is also unnecessary because in a number of the standard cases for the list the discriminatee can be both past- and future-responsible for possessing the trait. Thus, discrimination on the basis of religion is one of the classic traits for any prohibited list, but consider the following case of a religious convert: Person A has consciously and voluntarily changed from Religion R_1 to Religion R_2, and when queried responds affirmatively (and plausibly) that should she desire to do so, she could change her religious affiliation again. She is therefore responsible for her current religion in both the past- and future-oriented sense, but this will, of course, still be her religion, and were she discriminated against on that basis she would enjoy the protection granted by

religion figuring on the standard prohibited list. Even gender, probably as solid a candidate for an involuntarily acquired and immutable trait as any on the standard prohibited list, is principally subject to choice after the advent of modern sex-change surgery. In summation: responsibility for possessing the trait simply does not fit the list. (cf. Halldenius, 2005, p.461; Wasserman, 1998, p.807) However, as noted initially, I do not think we should consider the lack of fit with the standard list a decisive argument.

A second problem for the argument, it might be suggested, is that it will appeal only to those who believe that the normative notions upon which it relies should play some role in our ethical deliberations, and further that they should have this determinate role in the conceptualisation of discrimination. The underlying notion is, I believe, luck-egalitarian.[21] Utilitarians and libertarians, to take just two of the most obvious candidates, are unlikely to be willing to grant the required credence to such notions to get the argument off the ground.[22] However, I do not think we should give too much weight to this objection. The intuition supporting

---

[21] Luck-egalitarianism is, I recognize, itself a complicated and contentious concept. In the present I assume simply that luck-egalitarianism means that it is morally bad when a person is disadvantaged through no choice of her own. This is, I take it, broad enough to be uncontroversial, while sufficient for present purposes.

[22] For very different reasons, of course. For the first because, roughly speaking, what matters is the end-state total amount of good, not how distribution of it occurs or the individual position of any one person. For the latter because, roughly speaking, no individual carries moral responsibility for the position of other individuals except in so far as she is personally and unjustly the cause thereof.

the argument will find favour enough with many, including non-philosophers and those of us who are still somewhat agnostic (or just plain confused) about fundamental moral principles, to be worth investigating. As such, let us grant for the sake of argument that some form of luck-egalitarianism could principally support the case for the group-criterion.

Note however, how quaint the idea that it *actually* does so really is: when, as in the example above, someone is refused a job because of possessing the trait "being black", is it really that this is unfair because she did not and cannot *choose* not to be black? Is it not the reasonableness, or rather unreasonableness, of the requirement, rather than the person's lack of responsibility for meeting it which is central to the issue? As Tom Campbell observes: "It is only when the unchangeable requirement is in fact irrelevant, or the right in question is too fundamental to be denied to any human being, that unchangeability begins to have moral bite. [...] But why should we have to change these aspects of ourselves? It is the propriety of the requirement, not simply the possibility of meeting it that is at issue." (Campbell, 1991, p.158-159) This is well illustrated by the fact that there seem to be plenty of situations in which it is perfectly reasonable to treat persons differently because of their possessing (or not possessing) trait X, no matter whether trait X is voluntarily acquired, optional or both. As Alan Wertheimer argues in a parallel discussion of how to assess reaction qualifications: "It is not obviously wrong to prefer the teacher without a foreign accent [who facilitates learning by being easier for the students to understand] or the left-handed pitcher

[who will fit best against the opposing baseball team], although accents and handedness are (relatively) uncontrollable."[23] (Wertheimer, 1983, p.103; cf. also Heinrichs, 2007, p.106) But if a trait's being both involuntarily acquired and immutable fails to establish the moral relevance to discrimination of that trait, in that these qualities neither intuitively explain the wrongness nor uncontroversially imply the wrongness of differential treatment on the basis of that trait, it fails to establish these same groups as inherently relevant.

One possible answer to this objection is that there could be more wrongs involved with discrimination, and that the wrongness of discriminating on the basis of a trait that the discriminatee is not responsible for possessing generates only a *pro tanto* reason. If so, it may well be that the wrongness of discrimination is outweighed by other reasons, and so perhaps the benefits to the team or the students is what makes the use of a left-handed pitcher and a teacher with a native accent allowable, despite the wrongness of discrimination. I am undecided about how successful such a rescue attempt could be, but take it that the plausibility of the argument

---

[23] Some traits, such as race, might rarely, if ever, be contextually relevant. But one example that will strike many of us as legitimate is for a movie-director to hire only actors who are physically similar to the historical characters they are meant to portray, including sharing the same race. While the idea of a black actor playing Napoleon Bonaparte in a historical drama is not untenable, it seems reasonable to allow movie-directors to prefer a white person for the role. Or to put it differently: the reason that it is normally discriminatory to treat black persons differently than white persons *because* of their blackness is not that it is irrelevant because it is race, but that race is normally irrelevant.

for inherently relevant groups is at the very least weakened at this point.

But a final and very serious problem remains: past-oriented responsibility is not capable of supporting the group-criterion, and limiting the implications of a luck-egalitarian principle to future-oriented responsibility appears arbitrary, nor is there any obvious way of bridging the gap between the two. To see how this problem arises, consider first the list of traits that would be generated by the past-responsibility-account. Any trait, which the person possessing it is not responsible for having acquired, will go on the list. However, this seems to exclude no traits at all. Surely, the list of the traits that a person can be responsible for acquiring is limited by human nature, resources and inventiveness, but it is difficult to imagine a trait that a person could not somehow involuntarily acquire.[24] And even worse, we would have not one, but as many lists as there are persons, given that each must correspond to the traits that person was or was not responsible for acquiring. The past-responsibility account seems to lead not so much in the direction of a prohibited list as in that of a wrong-making

---

[24] At least setting aside traits that are, by their very definition, voluntarily acquired, such as the trait of "having voluntarily decided to do X". Nor are such beyond the scope of realistic cases – consider e.g. discrimination against those who voluntarily join the armed forces in a system composed of both conscripted and volunteers. Even so, I believe that the number of exclusions are sufficiently few as to make the resulting list implausible. As above, I would not want to claim that this "lack of fit" constitutes a definitive argument against a suitably strong account, but it does lend some weight to our concerns about basing the group-criterion on past responsibility.

principle, precisely because responsibility for acquiring a trait, on whatever account of responsibility we choose to apply, relates not to a quality of the trait but to the history of the person possessing it. As such, the past responsibility account cannot generate a stable list of traits, and cannot support the group-criterion.

The future-responsibility account could still support the group-criterion, however, and fits better with the traditional legal concern for immutability in any case. Even if there are potentially an unlimited number of immutable traits, there are also traits which will be excluded from the list, so the prohibited list will serve some purpose. But why should we accept a limitation to future-responsibility? Certainly, standard accounts of luck-egalitarianism would if anything tend to emphasize past-responsibility. Unless we have good reason to limit the scope of luck-egalitarian principles to future-oriented responsibility, the limitation is arbitrary. But such reasons not only have not been advanced, but it seems very difficult to imagine what they could be. We might even grant, as Larry Alexander argues, that the agent's lack of responsibility is an argument against immutable traits in the context of discrimination based on bias, i.e. the assertion that a person should be treated differently because she is of superior or inferior moral worth: "If discrimination is based on judgments of relative moral worth, then, of course, we good Kantians are likely to reject such judgments if they are based on immutable characteristics. Moral worth must be based on moral choices, not on physical characteristics or even character traits to the extent that such traits are not just proxies for the prior moral choices that formed them." (Alexander, 1992,

p.200) But this, as Alexander himself points out, does not support the argument for immutability, because it only highlights what we already knew: that moral worth can attach only, if at all, to an agent's choices. Nor does it exhaust our concerns for discrimination, which as Alexander recognizes is frequently based on other grounds than bias.

The fundamental problem, then, is that no matter the specific account of responsibility we adopt, it will tend towards a wrong-making principle rather than a trait-selecting principle and therefore will not fit something like immutability. To draw the distinction instead around immutability, we need a further argument, but it is not easy to see what one could look like. It is possible, of course, that one could be produced. But given the other difficulties the argument for inherently relevant groups faces, I think that until one is actually provided we can justifiably conclude that the argument does not seem capable of supporting the group-criterion.

Contextually relevant groups

The alternative way of justifying the group-criterion is to focus instead on contextually relevant groups, that is, to specify which groups should be the focus of discrimination relative to the current socio-historic circumstances, rather than relative to the innate and universally applicable qualities of traits.

Establishing the prohibited list within this approach relies on the connection between the traits and contextually determined social

circumstances or identity. The pertinent groups are groups with a history of past discrimination (Hellman, 2008), or groups that are, as Lippert-Rasmussen has aptly put it, "socially salient" in that: "… perceived membership of it is important to the structure of social interactions across a wide range of social contexts. Having green eyes is obviously irrelevant in almost any social context, whereas an individual's apparent sex, race, or religion affect social interactions in many social contexts." (Lippert-Rasmussen, 2006a, p.169) The focus is thus on a relatively well-defined set of traits, which are constitutive of groups that stand out because of their socially and historically specific group-identity. Identity can serve to do so both by internally constituting the group, through members' identifying with each other and any common problems, and by externally constituting the group, that is making it both possible and likely for other agents to act towards (and to *have* acted towards, in the past) members of the group *as* members of the group.

However, it still remains a pressing question why we should focus our attention on select contextually relevant groups. Why should discrimination be restricted to groups that are socially salient and exclude idiosyncratic acts? These questions are important, because it is too often at this point that the group-criterion ends up being implicitly endorsed, as when Wasserman highlights the social significance and importance to self-identity of the groups that "can be subject to discrimination", and then laments: "But while these features help identify the groups that can be subject to discrimination, they leave open the question of why it is especially

objectionable to take account of membership in such groups in denying a benefit or imposing a burden." (Wasserman, 1998, p.807) This strikes me as an at least partially erroneous diagnosis of the problem. Indeed, the most that can be said is that these features help identify the groups *that are commonly assumed to be relevant for discrimination*. I agree with Wasserman that identity does not immediately help us determine why discrimination is wrong, i.e. how it differs from differential treatment, but he seems to me to fail to raise the pertinent question. If identity did indeed serve to distinguish differential treatment from discrimination, then it would justify the group-criterion too: discrimination would be discrimination *iff* it was directed towards a socially salient group. But since this remains to be shown, then we must question not only what makes discrimination wrong – as is Wasserman's concern – but also what can justify the group-criterion.

Certainly there are good reasons to focus on groups with a shared identity from a practical point of view. Differential treatment involving members of these groups is likely to be where discrimination is most widespread and most invidious. But even if this makes it perfectly sensible to give special attention and more careful scrutiny to differential treatment on the basis of such group-membership, it hardly justifies applying the group-criterion to the definition, just as e.g. the fact that we should probably focus police resources on patrolling high-crime areas does not mean that our concept of 'causing harm' should include a restriction to make it apply only in those high-risk areas where physical violence is most likely to occur.

The best argument for the special relevance of socially salient groups seems to be that they are subject to harms which do not manifest to the same degree in idiosyncratic acts of wrongful differential treatment, i.e. what I shall refer to as the "additional harms-argument". Why do practices directed at socially salient groups cause more harm than idiosyncratic acts? Wassermann suggests several factors that may "contribute to the moral onus of taking group membership into account". (Wasserman, 1998, p.808) Let me mention just two: cumulative harm to individuals and social disharmony.

The cumulative harm of practices concerns the increased weight of burdens added to the already deprived: "If members of certain groups have been subject to worse treatment in a wide array of circumstances, it adds to the imbalance to disadvantage them on the basis of group membership. That effect will be amplified if members of those groups have a heightened concern about the treatment of other members, or about the fact that the adverse treatment they suffer is based on their membership." (Wasserman, 1998, p.807-808) Similarly, the additional harm of social disharmony relies on the preponderance of a particular form of discrimination in combination with the possibility of self-identification of the discriminatee with the discriminated group. In Larry Alexander's formulation: "One idiosyncratic use of a particular trait by a single discriminator is unlikely to affect the perception by members of the group defined by that trait of their general likelihood of obtaining positions and goods. For instance, if a particular employer wants his employees to have red hair, this

is unlikely to affect brunettes' and blondes' perception of their life prospects and thus their motivation and development of talents. On the other hand, if many discriminators use the same trait to exclude, motivational and psychic effects are more likely to occur, especially if many people perceive their personal identity largely in terms of possession of that trait." (Alexander, 1992, p.198) The costs of this effect must be counted both in terms of the stigmatization experienced by the discriminatees and by the loss of opportunities and productivity experienced by society as a whole, which results from the stigmatization, divisions and loss of confidence engendered.

The basic thrust of the additional-harms argument is therefore that discrimination must consist of actions that are capable of being social practices. An individual act of disadvantageous differentiation, while perhaps lamentable, can not amount to discrimination because it is incapable of causing the additional harms created by practices of disadvantageous differentiation. And since any and only socially salient groups can be the object of practices – the two are coextensive because any group that can be subject to a widespread set of similar responses, i.e. a practice, must be socially salient – social salience supports the group-criterion.[25]

---

[25] I intend here to include only realistic scenarios, although I admit that it is strictly speaking theoretically possible that a great number of people could simultaneously decide to apply idiosyncratic discrimination on the basis of the same non-salient trait. Some might hold this against the argument for

However, promising this might initially appear, I believe that in this form the additional-harms argument is unable to support the group-criterion. To see why, consider first that it seems obvious that harm can be done through idiosyncratic acts of discrimination. Indeed, proponents of the argument will typically admit as much, claiming only that additional harms occur in cases of discrimination against contextually relevant groups. But, if the additional-harms argument is an argument about *additional* harms, then it requires a threshold-argument to support the distinction involved in the group-criterion. That is, it needs to argue that there is a moral threshold at a certain level of harm, where the quality of the action changes, morally speaking, and that only discrimination against socially salient groups will cause sufficient harm to cross the threshold. This poses two challenges for it. Firstly, we will require arguments for why we should assume that only discrimination on the basis of socially salient traits will ever cross the threshold. This looks odd, for surely, we can imagine some act of idiosyncratic discrimination that caused more harm than standard acts of discrimination directed against members of a salient group? We can hypothesize very unlikely circumstances to produce an exceedingly great harm in a case of idiosyncratic discrimination if necessary, such as the complete lack of self-confidence and emotional vulnerability of the discriminatee, the fact that the discrimination and the trait in question is blatantly

---

contextually relevant groups, but I am willing to grant proponents that we should restrict our concerns to realistic scenarios in this manner. As such, I set it aside.

advertised, and the dramatic impact of this particular case of discrimination on the wellbeing and life prospects of the discriminatee, etc. The claim that in no hypothetical case will the harm of idiosyncratic discrimination exceed that of a reasonably defined threshold met by standard acts of discrimination against a socially salient group strikes me as utterly implausible. But the second requirement strikes me as even harder to fulfil, because the additional-harms argument also needs to provide reasons for why we should adopt a threshold in the first place. This requirement is similar to but importantly different from arguments for limiting legal prohibition to offences that cross a certain harm-threshold, and much harder to provide I believe. Indeed, I am hard pressed to imagine what such an argument would look like, and certainly, none has so far been provided.

The concern with weighing harm above might be taken to illustrate the consequentialist character of these arguments, but similar problems will afflict deontological accounts based on disrespect. Take that of Paul Woodruff as an example. On his account: "…an act of discrimination is wrong when it is wrong not simply because it is discriminatory, but because it is part of a pattern of discrimination that is wrong. A pattern of discrimination is wrong when it makes membership in a group burdensome by unfairly reducing the respect in which the group is held. It may accomplish this, for example, by making group membership a prima facie reason for failure. One act of discrimination cannot do that. If an applicant fails at one bank because of his race, and at other banks for other reasons, his race is not the reason for his

unemployment, and his failure is not an insult to his race. Discriminating, like walking on the grass, is to be judged with reference to how much of it is being done. Walking on the grass is harmful only if enough people are in the habit of doing it to ruin the grass. So it is with walking over the feelings of a group." (Woodruff, 1976, p.159) Here too, it seems clear that there is an implicit and implausible threshold-argument. After all, contrary to what Woodruff claims, the discriminatee's race will be the reason why she is unemployed, at the very least for the period of time between being denied that particular job and applying for the next, and so even an idiosyncratic act of disadvantageous differentiation will have some impact on respect, however miniscule. To deny this strikes me as a case of what Derek Parfit has labelled "the fifth mistake in moral mathematics": the fact that an individual act causes only a very small amount of harm does not mean that the act does not cause harm, and therefore does not mean the act is not (very slightly) wrong. (Parfit, 1984, p.75-82) On top of this, if Woodruff means to hold that there is not a continuous spectrum of greater and lesser but qualitatively identical wrongs, this seems to leave him beholden to the strange view that there is some point where adding one extra idiosyncratic act of disadvantageous differentiation directed at a particular non-salient trait somehow transmutes the collection of those acts into discriminatory acts proper, an act of ethical alchemy that I consider highly dubious.[26]

---

[26] Some proponents of the disrespect-account who focus on contextually

Failing this, the most that the additional-harms argument can say is that there are some forms of discrimination that are frequently worse than others, specifically that the forms of discrimination that focus on socially salient groups are for that reason more likely to cause harm, and likely to cause more serious harm, than those which are idiosyncratic. In practice there will be something like a graded scale of offences where some idiosyncratic acts of discrimination are in fact worse than some acts of discrimination directed at a member of a socially salient group; the overlap might be small, and even non-existent in some contexts, but this provides pragmatic reasons rather than conceptual ones for giving greater attention to socially salient groups. Contextually relevant groups, it therefore seems, cannot support the group-criterion.

Consider one last possible line of defence for the group-criterion. We have concluded that although there is no sharp normative

---

relevant groups recognize this. In a move parallel to Arneson's Deborah Hellman considers and initially apparently supports the idea of a group-criterion before eventually dismissing it, because of the greater potential for demeaning groups with a "history of mistreatment or current social disadvantage" (HSD): "Distinguishing on the basis of HSD traits may be morally different than doing so on the basis of non-HSD traits because the former reinforces or entrenches the caste-like aspects of our society. Laws that disadvantage groups without a social identity – people whose last names begin with A, for example – cannot reinforce a caste as there is no such social group whose status can be harmed or reinforced." (Hellman, 2008, p.22) She finally concludes, however, that: "Drawing distinctions on the basis of HSD traits has more potential to demean because of the social significance of such distinctions. But as I explained above, not all distinction-drawing on the basis of HSD traits is demeaning, as other aspects of the situation also affects whether one demeans. Moreover, categorizing on the basis of non-HSD traits can also demean; however, more contextual factors are required for this to be the case." (Hellman, 2008, p.28-29)

distinction between discrimination against contextually relevant groups and idiosyncratic discrimination, there is a difference of tendency between the two. Might this difference be sufficient to justify a weaker form of the group-criterion on pragmatic grounds? That is, might it be sufficient justification for including the criterion that usage thereby would track the tendency for the first to be worse than the second? Proponents of modelling the definition on common usage need not, I take it, be committed to the view that similar differential treatment directed against non-salient groups is necessarily morally different. Lippert-Rasmussen speaks in favour of this when he explains that: "An employer might be more inclined to hire applicants with green, rather than brown or blue, eyes. This idiosyncrasy might not amount to discrimination in the sense that interests us here, even though, obviously, the employer differentiates between different applicants. This is not to deny that such idiosyncrasies can be as bad as, and reflect as corrupted a character as, genuinely discriminatory acts. *It is just that in the great majority of cases they will not seriously harm the disadvantaged party, precisely because of their idiosyncratic nature.*"[27] (Lippert-Rasmussen, 2006a, p.169, my emphasis) And recall that although I expressed some reservation about tying the definition too closely to common usage, my main concern was that including

---

[27] Note that, as such, Lippert-Rasmussen cannot be said to employ a group-criterion in the sense I have discussed, because strictly speaking the restriction to 'socially salient groups' is conventional and tendential, rather than suggestive of the sharply delineated moral distinction required by the criterion. I owe thanks to Lippert-Rasmussen for helping me clear up my initial confusion on this point.

a distinction that was mere convention and tracked no interesting differences between the phenomena included and excluded was arbitrary and potentially misleading. This concern, it seems, might be alleviated when it emerges that there is in fact a difference, even if only of tendency, which the distinction tracks. Here, I confess to being less certain. It seems to me that as much may be lost by excluding phenomena that might warrant our consideration as instances of discrimination, as will be gained by focusing our attention on those cases likely to be the worst. But as we are now dealing with pragmatic grounds for choosing a convention, the issue is not one that need be resolved here; the group-criterion in the form I have sought to challenge remains untenable, even should we eventually decide that pragmatic reasons do speak in favour of a suitably circumscribed pragmatic group-criterion.

Conceptual and practical implications

As I have attempted to show above, it is difficult to come up with convincing arguments for the group-criterion. But what happens then, if we abandon the group-criterion and construe discrimination as possible along any trait? Perhaps the most immediate and important implication of abandoning the group-criterion is that predetermined groups plainly can no longer serve to explain the wrongness of discrimination. If there is no prohibited list, then there is no reason to discount idiosyncratic cases of discrimination. They are morally wrong regardless, for whatever reason discrimination more generally happens to be wrong. Thus, on e.g. Arneson's disrespect-account of discrimination: "If you discriminate against persons with large

earlobes on the basis of unwarranted animus or prejudice, this is wrongful discrimination, even if there is no history of wrongful treatment on this basis and no likelihood that you are setting a trend." (Arneson, 2006) This need not be considered a particularly troublesome conclusion. Even though there is no consensus on what does make discrimination wrong when it is morally wrong, the notion of a group-criterion tends to serve as an appendage rather than a central explanatory feature in the most recent literature on the topic. It simply means that there is no reason why we should consider discrimination against e.g. blacks or women to be worse than other forms of discrimination *per se*, and that we will need to look elsewhere for an account of what is wrong with discrimination.[28]

Having said that, we should recognize and retain the important insight of the additional harms-argument: that discrimination as a social phenomenon normally becomes the more grievous the more frequently and the more widely it targets a particular trait. A person who is discriminated against, e.g. as an applicant who is wrongfully denied a job, has suffered a wrong, no matter whether the act of discrimination is idiosyncratic (i.e. applies an unusual distinction) or an instance of a practice (i.e. applies a socially salient distinction). But she will be affected worse the more common this

---

[28] As I mentioned initially this is a contentious issue (See note 12 above). However, as none of the three approached mentioned there rely centrally the group-criterion, I take it that the loss of the group-criterion need not cause particular consternation.

form of discrimination is in society, both because the harm to her self-esteem is likely to be greater and because of her diminished prospects. If it is idiosyncratic her overall situation on the labour market is unlikely to be affected, but if it is widespread her chances of finding a job at all may be severely hampered, with all the losses of goods that this entails (economic, status, dignity, etc.).

A second and more dramatic conceptual implication is the way abandoning the group-criterion forces us to rethink indirect discrimination. On at least some accounts, indirect discrimination against group A is understood to occur when group A, which is on the prohibited list, is disparately affected by the non-discriminatory differential treatment of group B, which is not on the prohibited list but with which group A is either partially or fully co-extensive. Or, in the words of the EU-initiative: "When an apparently neutral specification, criteria or practice would disadvantage people on the grounds of racial or ethnic origin, religion or belief, disability, age or sexual orientation..." (Campaign, 2003) So a work-place, for example, which refuses to hire workers below 175cm of height is, on the conventional view, not engaging in direct discrimination against persons with a certain height, because this is not a relevant group. But it will probably be engaging in indirect discrimination against women, because many more women than men are members of this group, and as such women are disproportionately disfavoured by the requirement. However, if women are no more – and no less – protected than any other group, then explaining indirect discrimination in these terms no longer seems to make sense. If we are to conclude that

the refusal to hire persons of below 175cm of height is a case of discrimination, then it must be either a case of direct discrimination against such persons, men included, but *only* such persons, and therefore not against women including those who meet the requirement (i.e. are 175cm or taller), or it must be a case of indirect discrimination under a different account of that concept. This implication, although not in itself an argument for the group-criterion (at least not one that can avoid the fallacy of appeal to consequences), might be another reason that the criterion is hard to shed.

Practical implications

Apart from these conceptual implications, we might well ask what effects, if any, abandoning the group-criterion should have on practice. The implications for the legal system are obviously important. As we have seen, the group-criterion appears in the EU anti-discrimination-initiative's definition, and indeed in all legal definitions that I am familiar with. This is hardly surprising given that legal protections are directed at addressing socio-historically specific instances of injustice, and the necessity of specifying legal prohibitions in terms unequivocal enough that they can form the basis of positive law.

But the question of how to transform our ethical obligations, in the shape of the anti-discrimination principle, into law is a complicated one that hinges, among other things, on whether one adopts a liberal stance that considers many forms of discrimination to fall below the threshold or outside the scope of harms that

would justify legal prohibition, or a more restrictive stance that considers the harms of discrimination grave enough for justifiable prohibition to be fairly encompassing (cf. e.g. Brest, 1976; Campbell, 1991; Cohen, 1994; Gardner, 1989; 1996; McCrudden, 1985; 1982; Waldron, 1985). Although, because of this complicated relationship between morality and law, the practical implications pose questions that deserve more extensive treatment than I am capable of giving them here, my feeling is that the ramifications of abandoning the group-criterion will be less dramatic than one might fear.

Whatever approach to prohibition one favours, the immediate implication of abandoning the group-criterion is likely to be the necessity of specifying in greater detail what is discriminatory and when. As wrongful discrimination does not simply follow a division along the traits that form the basis of the discrimination, and as a general prohibition on "discrimination" would, given the lack of consensus, beg the question of when differential treatment is or is not morally wrong, we may need to focus legislation on those situations in which specific forms of differential treatment is uncontroversially wrong. This is not necessarily an unwelcome consequence. One unfortunate effect, it seems to me, of the current use of the group-criterion in legislation is that it often serves to gloss over underlying and controversial obscurities. I basically agree with Richard Arneson, albeit for different reasons, when he suggests that: "…the idea of wrongful discrimination is not going to do much heavy lifting for the task of determining what social justice requires with respect to policies for dealing with

suspect classifications." (Arneson, 2006, p.796) We will have to look carefully at the specific context and the specific trait in question instead. And it might turn out to be both more just and more expedient to make prohibitions more specific, seeing as how blanket prohibitions against any form of discrimination involving groups on a prohibited list are not, as I have tried to show, normatively sustainable.

Many prohibitions against discrimination in more specific areas of law in fact already employ relatively detailed specifications, as in the case of employment law which frequently includes a qualification prohibiting discrimination on the basis of traits on the prohibited list, except where such traits are contextually relevant to the function of the job.[29] This renews the important discussion of what constitutes such relevance, embodied e.g. in the notoriously controversial principle of "business necessity" in the US Civil Rights Act of 1964, (cf. Connolly, 2006, p.127-132, 153-157) including whether and if so when reaction qualifications can form the basis of legitimate discrimination, but is, I believe, a step in the right direction. We will need to resolve such difficult questions not by glossing over, but by careful deliberation.

One way to adapt such prohibitions to life without the group-criterion would be to simply leave out the reference to specific

---

[29] For an interesting example of a law that attempts such detailed specifications see the UK Equality Act 2010, as well as its predecessor of 2006 (Equality Act 2010, 2010; Equality Act 2006, 2006).

groups, resulting in a general prohibition against arbitrary discrimination, or to put it the other way round: a requirement that differential treatment be justifiable with reference to the relevance of the distinguishing trait in the context. The challenging part of this solution is that it confronts directly the vagueness, or even the inherent contradictions, of our understandings about what constitutes the morally right act in many situations. Thus, for e.g. employment discrimination, to leave out the group-criterion would mean to prohibit *all* instances of treating employees differently that advantage some and disadvantage others, except where such differentiation can be plausibly argued to be based on a trait which has moral relevance to the employment situation. Not only does this bring back the difficulty of determining contextual relevance with a vengeance, it also shifts the burden of proof by introducing a strong meritocratic norm which flies in the face of the powers that are generally assumed to be the employer's prerogative, e.g. to hire and fire workers at her discretion. While I am not sufficiently attached to the liberal conception of labour market ethics to consider this an inconceivable conclusion – so much the worse for employers' prerogatives we might say – it is certainly one that is so far removed from our current practices that it would require careful investigation before we adopt it.

Another option might be to grant the importance of the additional harms-argument in this context. Given a liberal approach to the criminalization of discrimination, the additional harms caused by discrimination against socially salient groups might be the condition that pushes the harm above the threshold and justifies

legal prohibition: "On this view, unlawful discriminatory classifications are those whose use gives rise to serious inequality arising from social prejudice." (Campbell, 1991, p.154) Thus, while it probably does not make sense to prohibit discrimination against e.g. women *tout court*, given the lack of a normative foundation for the group-criterion, it may make sense to employ a narrow prohibition of discrimination against women which does not include arbitrary discrimination broadly speaking, in specifically defined situations such as in those workplace situations where gender does not constitute a relatively obvious and uncontroversially relevant trait.

...but some groups are more vulnerable than others

In the course of this article I have attempted to show that what I have dubbed the group-criterion, i.e. the notion that there are certain, predetermined groups which are relevant for the concept of discrimination, and that by implication all others are not, must be considered untenable. I have examined two approaches to justifying the criterion, one based on inherently relevant groups and one based on contextually relevant groups, and found both incapable of doing so, rendering the distinction ultimately arbitrary. It remains possible of course, although I think it unlikely, that unexplored arguments could support the criterion. Until they appear, we must examine instead what will happen if we drop it from the definition of discrimination.

I have tried in the last section of the article to sketch what I take to be a number of possible implications of abandoning the group-

criterion, including the consequences for the idea of indirect discrimination and discrimination law. Doing so clarifies the concept in important respects, but raises new questions and complications in others. I have not attempted to resolve these challenges here, but have suggested that even without the group-criterion there may be good reasons why positive law should direct its definitions to the protection of particular, predetermined groups. I certainly do not think that the conceptual clarification I have advocated can serve to argue against using such categories as a legal tool, nor against discrimination law more generally. If anything, the implications of abandoning the group-criterion might be an expansion of our legal prohibitions to include cases of discrimination which are currently not conceived in those terms.

It is worth stressing the following point, that my argument against the group-criterion means neither, of course, that all kinds of discrimination are equal in practice nor that discrimination is not morally wrong. There are some forms of discrimination – racism, misogyny, islamophobia and homophobia to name a few obviously important cases – that have been and continue to be invidious and shameful social problems in a way that the various hypothetical cases of idiosyncratic discrimination I have discussed are not. As such, they undoubtedly merit much greater attention and concern. Only, I would say, for pragmatic rather than conceptual reasons.

What clarifying our concepts in the way that I have attempted should do ideally, I suppose, is to heighten our awareness of the obligations enshrined in the principle of non-discrimination as part of a larger set of moral norms, which are perhaps both more

extensive and more demanding than we tend to imagine. Discarding the group-criterion, with its insistence that some groups are more equal than others, does not mean giving up the egalitarian ideal inherent to the principle of non-discrimination, but rather forces us to think more carefully about what it means for persons to be equal, and what challenges this poses for us as moral beings.

Stealing Bread and Sleeping Beneath Bridges - Disparate Impact, Indirect and Negative Discrimination[†]

Introduction

Most persons today will probably agree that discrimination is morally wrong and a social evil that should be opposed. It is typically less clear, even to those who feel strongly about discrimination in this way, exactly how to define it, what the moral wrongness consists in and as a result thereof how to asses non-standard cases of discrimination. Consider the following example of what I believe is a fairly uncontroversial case of discrimination:

> *The Mogul Misogynist.* A factory is purchased by a wealthy financier, who reviews the personnel records, and finds that about half the workers are male, and half female. The mogul then proceeds to order the female workers fired and replaced with male workers, because e.g. he holds mistaken beliefs about female workers' efficiency, has a brute preference for male workers or considers women morally unsuitable for factory work.

The actions of *the Mogul Misogynist* are an example of what is normally called direct discrimination. Direct discrimination against a person can be roughly defined as occurring when:

1) an agent differentially treats persons with trait
T (T-persons),
2) the differential treatment is suitably explained
by T-persons possessing trait T,[1]
3) the differential treatment negatively affects T-
persons, and
4) the differential treatment is morally wrong.[2]

Although surprisingly underdeveloped in the philosophical literature, given the prominent role that discussions of discrimination plays in public debate and the strongly condemnatory meaning that the concept connotes, recent years have seen some fruitful discussions that have both clarified discrimination conceptually and explored different accounts of what makes discrimination morally wrong, when and if it is morally wrong. (Radcliffe Richards, 2000; Halldenius, 2005;

---

[1] Note that the notion of "suitably explained" must be understood as something other or more than simply intentionality. Otherwise, we should have to say that a Mogul who holds prejudiced false beliefs about women's inefficiency discriminates against the inefficient. In this case it seems to me sufficient as a suitable explanation to say that it is counterfactually true that were the discriminatees not women, they would not have been subjected to differential treatment. I return to this issue in the section on "Intentionality, the irrelevance of irrelevance and protected groups" below.

[2] Note that for the purposes of this article I thus limit the concept of discrimination to morally wrongful cases of differential treatment, in line, I think, with increasingly common usage. Much of the literature instead applies a morally neutral definition of discrimination and applies the epithet "morally wrongful". Note also that while this is a very rough definition I believe it is suitable for our purposes here. For the more precise and detailed definition on which this is based, including discussions of what it means to negatively affect T-persons, cf. (Lippert-Rasmussen, 2006a; 2007a)

Arneson, 2006; Edmonds, 2006; Lippert-Rasmussen, 2006a; 2007a; Heinrichs, 2007; Hellman, 2008; Lippert-Rasmussen, 2008; Moreau, 2010; Altman, 2011; Segall, Forthcoming) Now consider a second variation:

> *The Mogul Misogynist 2.* The financier purchases the factory, and intends to fire the female employees for the same reason(s) as in the first variation, but realising that this would violate anti-discrimination laws he instead orders two new assembly belts installed that will somewhat increase productivity. The height of the assembly belts require workers to be at least 175cm tall, allowing him to replace employees 174cm or shorter, which results in all of the female (and some of the male) employees being fired.[3]

---

[3] Of course these figures are fictions and could be made up to look any way we want to make the case work, but the assumption that all female employees could lose their jobs used in this scenario is not implausible. Suppose average male height is 180cm, and average female height is 165cm. These are relatively close to figures for young adults in Northern Europe and North-America according to Wikipedia. (http://en.wikipedia.org/wiki/Human_height) Given a standard deviation of ca. 7cm (again according to Wikipedia) this means that only ca. 23,9% men will be shorter than 175cm, while ca. 92,4% of women will be shorter than 175cm; on average female employees will constitute ca. 79,4% of those fired. While it is unlikely that any sizeable factory will not employ at least one woman of sufficient height to retain her job, a factory with e.g. 30 employees (half male, half female) has a ca. 30,5% chance of firing every single woman upon installing the assembly belts.

The facts of the case are not much altered, but what we have in the second variation is not a case of direct discrimination as we defined it above, because the trait that satisfies conditions 2 (link between T and treatment) and 3 (link between T and negative effect) is different from the one that satisfies condition 1 (differential treatment between T and non-T). The second variation would, if interpreted as a form of direct discrimination compatible with this definition, be a case of discrimination against "people shorter than 175cm", which while perhaps closer to the truth than it seems, is not, I believe, how most persons would intuitively describe the situation.

What we have here is rather a case of what is commonly described as indirect discrimination. Conventionally, and here in the words of the European Commission's Department of Employment, Social Affairs and Equal Opportunities, indirect discrimination is understood to occur: "When an apparently neutral specification, criteria or practice would disadvantage people on the grounds of racial or ethnic origin, religion or belief, disability, age or sexual orientation…"[4] (Campaign, 2003)

So a work-place which refuses to hire workers below 175cm of height is, on the conventional view, not engaging in direct discrimination against persons of 175cm or lower height, because

---

[4] Note that, somewhat surprisingly, and inconveniently given the example I have presented in *Mogul Misogynist*, the list does not contain gender. Gender is, however, a standard trait on every other such list that I am familiar with.

this is not a relevant group. But it will probably be engaging in indirect discrimination against women, because many more women than men are members of this group, and as such women are disproportionately disadvantaged by the requirement. I shall want to adopt a slightly broader definition, for reasons to which I will return, which holds simply that indirect discrimination against T-persons occurs as a consequence of non-discriminatory differential treatment of T_1-persons, with which group T-persons is either partially or fully co-extensive.

Indirect discrimination, then, can be preliminarily defined as occurring when:

> 1) an agent differentially treats persons with trait T,
> 2) the differential treatment is suitably explained by T-persons possessing trait T_1,
> 3) the differential treatment negatively affects T_1-persons,
> 4) because (at least some) T_1-persons constitute a disparately affected sub-set of T-persons,[5] and
> 5) the differential treatment is morally wrong.

---

[5] Note that as brought out more clearly here than in the previous definition, the suitable explanation in condition 2 is not related to intention, but merely a requirement of a causal connection between the possession of trait T and the differential treatment. In this example the financier's intentions concern T_1-persons.

In the literature, comparatively little attention has been granted to the notion of indirect discrimination, even though such cases are potentially more common than cases of direct discrimination, simply because any case of differential treatment can involve multiple forms of indirect discrimination. In an influential article on the topic Kasper Lippert-Rasmussen specifies only that: "Indirect discrimination occurs whenever an individual, institution, or practice acts in such a way that the interests of some individuals are systematically favoured and yet this does not involve direct discrimination." (Lippert-Rasmussen, 2006a, p.170-171) Similarly, in a recent article on the grounds for legal prohibition of indirect discrimination, Oran Doyle takes as his starting point simply the definition that "'Indirect discrimination' connotes a measure that does not on its face distinguish between class A and class B but which, for some related reason, is nevertheless considered troubling." (Doyle, 2007, p.538) What I want to argue in the following is that the distinction is vaguer and more complicated than appears as well as being less interesting than is often assumed, in that it misidentifies the features of situations such as *Mogul Misogynist* which is relevantly different between variations 1 and 2. Further, that properly understanding the relevant difference points to two important conclusions: first, it provides indirect support for a harm-based account of discrimination in that the central feature rests on an underlying concern for how the welfare of persons subject to discrimination is affected, and secondly, it suggests that our moral obligations qua non-discrimination are substantially broader than is frequently assumed.

Intentionality, the irrelevance of irrelevance and protected groups

The first question we might ask of the definition suggested above is what more precisely could constitute a "suitable explanation" qua condition 2. That some form of this condition is necessary in cases of direct discrimination seems obvious, since we would not want to classify the giving of a lottery-prize to a man as a case of differential treatment "for men" or "against women". The connection between the differential treatment of this person and his gender is completely arbitrary. Even were we to draw a lottery-ticket for every human being alive, and were the lottery miraculously to come out so that every single man won a prize and no woman did, it would still seem decidedly strange to call this a case of differential treatment against *women*, which is not to say that it might not be an undesirable outcome or a problematic procedure for other reasons. Although men and women are differentially treated, and the outcome of the treatment is favourable to men and disfavourable to women, the treatment does not track the possession of gender in the required sense. What is missing is some form of non-random link between gender and the differential treatment.

An obvious suggestion might be to let intentionality fulfil the function of suitable explanation, so that differential treatment of T-persons and non-T-persons can only constitute discrimination if

it is in some sense *directed* towards treating T-persons and non-T-persons differently *because* they posses or do not possess T.[6] Additionally, it might seem that the different ways that intentions manifest in direct and indirect discrimination highlights an important difference between them. Thus, in *Mogul Misogynist 1*, the agent acts directly upon his intention, while in *Mogul Misogynist 2* he realizes it circumspectly, at once cloaking his misogyny and inserting an intermediate step in the process. Furthermore, the intentional focus will perhaps be an attractive feature to those who believe that the wrongness of discrimination is (at least partly) accounted for in terms of the mental-states of the agent.[7] (Alexander, 1992; Arneson, 2006; Glasgow, 2009) I believe, however, that it fails to include plausible cases of indirect discrimination.[8] Consider this third variation:

> *The Mogul Misogynist 3.* The financier purchases the factory, but has no intention of firing the female employees, nor does he hold the bias, preference or belief of the first two variations. He orders two new assembly belts installed that

---

[6] I use this slightly broad formulation deliberately, because we may want to talk of e.g. laws or institutions discriminating despite their lack of the ability to form intentional mental states.

[7] Such accounts typically explain the wrongness in terms of disrespect, but there are disrespect-accounts which do not rely on the mental-states of the agents. Cf. (Hellman, 2008)

[8] Indeed some might want to hold the stronger claim that indirect discrimination is unintentional by definition. I do not find this a particularly obvious claim, but if one accepts it, it simply strengthens my argument here. As such, I set it aside for present purposes.

will somewhat increase productivity. When informed that the height of the assembly belts as standardly installed will require workers to be at least 175cm tall, he responds by firing and replacing those 174cm or shorter, which results in all of the female employees being dismissed.[9]

In this third variation, the mental state, including the intentions of the mogul, has no immediate bearing on the differential treatment of women in the way of variations one or two, nor does it arguably explain the wrongness of discrimination, because the mogul has no intention of discriminating against women and holds no bias, dispreference or mistaken beliefs about those affected. Any discrimination against women occurring in variation three is unintentional on his behalf. Nevertheless, we might still want to say both that what happens is morally wrong and that it is not a case of direct discrimination against the less-than-175cm'ers but a case of indirect discrimination against women.[10] If this is the case, intentionality will not do.[11]

---

[9] Note that to make the examples intuitively easier to grasp we can assume that there is no particular reason why one of the belts could not have been set to the standard height and the other to a lower height appropriate for (most) women. This simply avoids inserting potentially disturbing factors such as the balancing of economic efficiency against the disparate impact, etc.

[10] This seems intuitively right to me, but the present context is not the place to develop the detailed arguments to support that intuition. In the following therefore I shall simply assume that variation 3 is a case of, or very likely to be a case of, morally wrongful discrimination, and that as such the mental-state

112

Note something else, however: in defining indirect discrimination we have introduced a different form of causal connection in the shape of the condition that concerns disparate impact. Here, it seems, the fact that women are differentially treated is suitably explained by the fact that women constitute a disparately affected subset of the less-than-175cm'ers.[12] This suggests that rather than supplementing it, condition 4 in the suggested definition (condition 3 in the revised version below) could *replace* condition 2, as the suitable explanation of how discrimination occurs, so that indirect discrimination occurs simply when:

> 1) an agent differentially treats persons with trait T,
>
> 2) the differential treatment negatively affects T_1-persons,
>
> 3) because (at least some) T_1-persons constitute a disparately affected sub-set of T-persons, and

---

principles reviewed above do not satisfactorily explain what is going wrong in the *Mogul Misogynist* variations.

[11] There are other, and more fundamental problems with both mental-state and disrespect-accounts of the wrongness of discrimination. I explore these at greater length in "Discrimination and Disrespect". (cf. also Lippert-Rasmussen, 2006a; Segall, Forthcoming)

[12] Two questions emerge at this point: first, whether we could allow their constituting such a subset to be random, in the way that my example of the lottery illustrated, or whether there needs to be a non-random connection between the two traits so that possessing one is non-arbitrarily connected to possessing the other; second, what type of disparate impact is required, as this could be disparate impact of kind or quantity, i.e. severity of impact or numbers affected. I shall return to these questions below, arguing that the link still needs to be non-random and that it must be disparate impact of kind, not quantity.

4) the differential treatment is morally wrong.

This raises a different challenge, because without intention to specify which group is targeted the definition now looks as if it might fail to give us the conceptual tools for deciding whether cases like *Mogul Misogynist 2* are instances of direct or indirect discrimination. While it seems natural to describe *Mogul Misogynist 2* as something other than a case of direct discrimination against persons lower than 175cm, why should this be so? We need to know what it is we are overlooking if we do.

Frequently, discrimination is conceived as inherently limited in scope in ways which might answer this question. Thus one tempting approach might be to explain the wrongness in terms of the irrelevance of the criterion used to differentially treat T-persons. What we may call the *relevance account* of indirect discrimination would thus hold that race, gender, religion, etc. are morally irrelevant traits, that differentially treating on the basis of irrelevant traits is inherently wrong, and that indirect discrimination occurs when apparently benign discrimination causes differential treatment on the basis of an irrelevant trait. The difference between direct and indirect discrimination therefore concerns whether or not the selection-criterion is one of these morally irrelevant traits.

Some of the temptation towards the *relevance account* may derive from our familiarity with legal practice, which in many countries allows indirect discrimination when the selection criterion is subject to "objective justification". (Connolly, 2006, p.153-170)

That is, e.g. employers are permitted to hold requirements that disparately impact a relevant group when these requirements can be shown to be pertinent to the job in question. Thus, even if more men than women have commercial driver's licences, a trucking company that requires applicants to hold such licences is not considered to indirectly discriminate against women despite the fact that women will be disproportionately prevented from applying.[13] Much in vogue in at least Anglo-Saxon legal systems since the famous *Griggs vs. Duke Power Company* US Supreme Court case, (Griggs vs Duke Power Co., 1971) the requirement of objective justification also has the advantage that it circumvents some of the problems attached to alternatives, such as the challenge of demonstrating discriminatory intent. (cf. e.g. McCrudden, 1982, p.329-336)

Whether or not objective justification approaches to indirect discrimination are the best way of framing legal protection is a separate matter that I shall not consider in the present context. But as an account of the morality of differential treatment, and as a means of distinguishing between direct and indirect discrimination, the focus on irrelevance runs into severe problems. First, in the definition of relevance, secondly, in explaining why treating persons differently for irrelevant reasons should in and of itself

---

[13] This is putting it too boldly, perhaps. There is a rich and complex body of laws on the objective justification defence in indirect discrimination cases, and in some scenarios it might not apply, even in such intuitively reasonable circumstances as above. I trust, however, that the general principle is clear from the example, and set added legal complications aside for present purposes.

make a moral difference, and finally in distinguishing cases involving more than one irrelevant trait.

First, it is far from easy to define the relevance involved in a limiting and non-question-begging manner. If we take it to be simply what the involved agents regard as being relevant then it seems clear that pretty much anything can be relevant given the suitable conditions and context of the agents. (cf. Wertheimer, 1983) On the other hand, if we take it to be a more substantive, morally circumscribed notion of relevance, then this raises the twin challenges of supporting this notion independently and avoiding that in doing so we make relevance superfluous, in which case we are no longer dealing with a relevance account.[14] Second, it is not inherently plausible that acting on the basis of morally irrelevant traits leads to moral wrong. Surely, if I am handing out candy, and give girls green-wrapped caramels, and boys yellow-wrapped caramels, then although my differential treatment is based on no relevant difference – suppose that I myself claim no relevant reason for the differential treatment – it seems intuitively absurd to say that I am engaging in a moral wrong. Third, the account would give us no reason to prefer describing *Mogul Misogynist 2* as a case of indirect discrimination against women rather than direct

---

[14] As I shall return to, this is in fact what I take to be the case. That is, the intuitions that support the relevance-account are best understood as really pointing towards an account that relies on a moralized conception of what is relevant, specifically a welfare-account. But then we are no longer dealing with something that is best conceived as a relevance account.

discrimination against persons below 175cm, because presumably being 174cm or shorter is a trait equally morally irrelevant.

A related but alternative way of explaining the intuitive difference at stake in indirect discrimination is by reference to the target group. Most legal and political conceptions of discrimination operate, as we saw in the example of the European Commission's definition above, with a list of protected groups which must not be the victims of discrimination. Indirect discrimination is then simply understood as discrimination which affects one of these groups indirectly. The *protected-groups account* would thus hold that the differential treatment in direct discrimination is based on a trait that is morally *relevant* in the sense that discriminating on the basis of it is inherently morally wrong, while differential treatment in indirect discrimination is based on a trait that is morally irrelevant, in the sense that discriminating on the basis of it is morally benign. However, indirect discrimination is morally wrongful, and distinct from properly benign differential treatment, because it disparately affects persons in a way which produces an outcome similar to that created by direct discrimination, i.e. is indirectly discrimination on the basis of a morally relevant (protected) trait.

The problem is that it is extraordinarily difficult to explain why a limited number of groups, and specifically those groups which traditionally figure on such lists, should be granted special moral status in the way which the account supposes. Being a woman, or black, or muslim, are traits that we rightly feel are under normal circumstances morally irrelevant – that after all is a premise of the standard argument against direct discrimination on the basis of

membership of these groups. For that very reason, there is on the face of it no more reason to worry about women or blacks or muslims being disparately affected by differential treatment of T-persons, than there is to worry about any other group based on morally irrelevant traits being disparately affected.[15] We rightly do not worry about whether a policy will or will not affect more people with three-syllable first names or two-syllable first names. In a classic article, Larry Alexander analyses the morality of discrimination primarily through the justifiability of the preferences involved, and suggests plausibly that, at least when the preferences are themselves benign, there is no intrinsic value to distributing the same (uneven) portions of goods among otherwise equal individuals one way or the other, no matter their race, gender, religion, etc. All such distributions will be equally good.[16] (Alexander, 1992, p.183-188; cf. also Lippert-Rasmussen, 2008)

---

[15] There are a number of possible arguments, all of which I believe fail. I discuss these at some length in "But Some Groups Are More Equal Than Others".

[16] Though I am sceptical of other parts of Alexander's analysis, this is a conclusion that I shall adopt as a premise in the following. Note, however, that some distributions may have instrumental value, because e.g. of the symbolic effects of distributions, just as individuals are unlikely to be otherwise equal. Differences in distribution of goods prior to the distribution in question will make a difference on any plausible account, just as prioritarians and egalitarians will want to take into account prior and resulting comparative welfare respectively. As it turns out, Alexander is moderately sceptical about putting the burden of distributing goods in the optimal way on moral agents, and certainly opposed to most forms of legal prohibitions. (cf. Alexander, 1992, p.194-198, 212-217) Here, although I think there are legitimate concerns about whether and in what shape to institutionalize such requirements in law, I am much less concerned than Alexander about the moral burdens placed on individuals.

However, if women are no more – and no less – protected than any other group, then resolving the issue of whether *Mogul Misogynist 2* is best understood as direct or indirect discrimination through its direct or indirect impact on protected groups no longer makes sense. Could the two concepts be not mutually exclusive, that is, could *Mogul Misogynist 2* be a case of both? Perhaps, but if so we are left with the tricky task of specifying how the moral wrongs involved relate to each other. Do both descriptions concern one act that is morally wrong for two reasons, two acts that are wrong for one reason, or two acts that are wrong for two reasons, and how exactly do we avoid double-counting given that most of those wronged will be subject to both direct and indirect discrimination? On top of this challenge, the idea flies in the face of our intuition that *Mogul Misogynist 2* ought to be about women, not the less-than-175cm'ers. We strongly feel that we should worry about women, blacks, muslims (and many other socially salient groups) in a way that makes it intuitively preferable to describe *Mogul Misogynist 2* as a form of indirect discrimination against women. Perhaps it is time to give more careful consideration to why. In doing so, I believe we can at the same time dispense with much of the obfuscation created by the concept of indirect discrimination.

What's so indirect about discrimination?

In the above, we first dispensed with intentionality and then investigated two common-sensical solutions for the challenge of providing a distinguishing criterion between cases of direct and indirect discrimination. As these failed, I briefly explored the

possibility that we might not need such a distinguishing criterion, and concluded that it did not look promising. A simpler solution, and one that better tracks our intuition that the indirect discrimination against women is more important than the hypothetical direct discrimination against the less-than-175cm'ers, might be to hold on to exclusivity, that is, to insist that *Mogul Misogynist 2* is either direct or indirect discrimination but not both, and to focus on the condition of moral wrongness as decisive of whether it is the first or the second.

Since we are now exploring the role played by moral wrongness in discrimination, I should emphasize that, although I do not want to assume a particular account of the moral wrongness of discrimination, I do assume, when I speak of discrimination being morally wrong, that it is only *pro tanto* wrong.[17] On most moral theories, e.g. consequentialism and threshold-deontology, there will be situations in which discrimination is all things considered permissible (or even required), despite its being all else equal impermissible. Or putting it differently: when comparing counterfactuals, the situation may be one where differential treatment is morally permissible, because the hypothetical situation where the agent does not differentiate leads to a sufficiently bad outcome, but it can remain *pro tanto* morally wrong, and hence discrimination, if we would be morally obliged to prefer a second

---

[17] I shall argue below that my conclusions in the following lend indirect support to the harm-account of the moral wrongness of discrimination, but this of course is reasoning the other direction.

counterfactual in which we could avoid both the act of differential treatment and the bad outcome which it prevents.

How might we argue for classifying *Mogul Misogynist 2* as indirect discrimination based on the condition of moral wrongness? The first point to notice is that while we failed to account for the difference by specifying morally irrelevant factors or protected groups, one factor which seems uncontroversially morally relevant to cases of discrimination is whether or not the differential treatment harms discriminatees. If for example the discrimination disadvantages the discriminatees by reducing the level of welfare that they would otherwise have enjoyed, then this makes a moral difference.[18] In effect, such welfare-reductions are causing harm. Very few, if any, plausible moral theories will hold that this is not at least one factor that bears on the moral permissibility of the action; pluralists, of course, will hold that other factors do too.

Let me give just three examples of how discrimination might cause harm by affecting the welfare of the discriminatee:[19] First, the

---

[18] This is somewhat of a simplification, because it avoids the question of what the relevant baseline for individuals' welfare is. If it is determined e.g. that the target of an action of disadvantageous differential treatment currently enjoys goods that although beneficial to her welfare ought properly to be otherwise distributed, then although her welfare may be reduced as compared to her level prior to the action, this reduction brings her closer to the level that she ought to enjoy, and is arguably not morally wrong, and thus not discrimination on the definition here adopted. (cf. Lippert-Rasmussen, 2007c) I set such complications aside for present purposes.

[19] Some might want to describe the harms of discrimination in terms other than the impact on welfare of the discriminatee, such as e.g. placing restrictions on the opportunities available to persons. (Moreau, 2010) I do not mean to argue

public differential treatment of the two groups may cause symbolic harm. That is, it may send a signal that discriminatees are in some respect inferior, which is at once insulting to discriminatees and potentially creates or reinforces prejudices about the target-group which are likely themselves to be at least instrumentally bad. Thus, the harm at stake is symbolic by means, not by nature, and very real. The risk of causing symbolic harm is particularly significant when the group in question is a socially salient group, and even more so when there is a history of prejudice and discrimination against the group.[20] Second, disparate impact may occur in the shape of inegalitarian distributions. The distribution of goods crucial to the welfare of discriminatees will matter on any account that considers global outcomes to be of moral importance, whether e.g. egalitarian, prioritarian, sufficientarian or maximising, due among other things to the diminishing marginal rate of utility. An uneven distribution or redistribution of goods may of course deny those who ought to have certain goods these goods, but even an equal redistribution of goods may also fail in this respect, if it does not remedy existing inequalities. Third, disparate impact may

---

for or against a particular conception of harm in the present. My point can, I believe, be carried through with suitable modification of the examples for alternative accounts of harm.

[20] This, I believe, is in fact one of the two main reasons that our intuitions tend to emphasize discrimination of protected groups, the other being the far greater likelihood that these groups will be subject to discrimination. One might add that such groups will typically be socio-economically vulnerable in a way which makes them susceptible to harms resulting from in-egalitarian distributions, but this is at once true of many groups not traditionally placed on the list of protected groups, and not true for at least some of the groups often placed there, so I take it to be less crucial in driving our intuitions.

occur as a consequence of the differing preferences of discriminatees. Clearly, if I serve sugar-rich food to diabetics or pork to Jews or Muslims, I am failing to take into account that these groups have dietary needs that are incompatible with my choice of cuisine. Note that inegalitarian distributions shares certain features with differing preferences; roughly, the difference between them consists in that the first may be said to concern the distribution of different *quantities* of goods, while the second concerns the distribution of different *qualities* of goods, where both quantity and quality is relative to the receiver. Note also that all three forms of disparate impact may of course occur simultaneously from a single act of discrimination.

The second point to notice is that disparate impact is inherently part of discrimination. It is because the treatment adversely affects one party, in some respect, and not the other that it is differential treatment, not necessarily because the agent acts in two different ways. Thus, while many of the most prominent examples of discrimination involve the agent acting in one way towards T-persons and acting in a different way towards non-T-persons, whether it be hiring men while firing women or stop-searching blacks while ignoring whites, this need not be the case. Differential treatment as the concept is standardly employed does not preclude the agent acting in a way that can relevantly be described as identical towards the two groups. As in the example of differing needs above, if I serve pork to both those who desire eating it and to those for whom eating pork is prohibited by a strongly felt religious taboo, then under one relevant description I am acting in

exactly the same way while serving both dishes. Of course, under a different description of my actions, one that takes into account certain morally relevant elements of the context, specifically the dietary needs of those to whom I serve food, I am doing two different things, i.e. serving edible and inedible food. But although I would not want to claim that one of these descriptions is more accurate than the other, it seems more natural to apply the first description, which I take to be sufficient to establish my point: that it is possible to differentially treat two groups while performing actions that we can naturally and accurately describe as being identical.

What we have then are two kinds of differential treatment, the first of which disadvantages T-persons because the actions of the agent towards T-persons are different, under the most natural action-description, and the other of which disadvantages T-persons because the actions of the agent are identical under the most natural action-description.[21] Central to our understanding differential treatment in this way, however, is that the discriminatees in the first case are identical in certain pertinent respects, a likeness which is not reflected in the different actions undertaken, while the discriminatees in the second case are

---

[21] Note that we can set aside cases where the differential treatment does not disadvantage T-persons (or non-T-persons) because such cases are by definition not discrimination *against* persons, although it could be conceived of as differentiating *between* persons. A standard example concerns gender-separated toilets, which seem to be harmless, or even decidedly advantageous, to members of both sexes.

different in certain pertinent respects, a difference which is not reflected in the identical actions undertaken. Let me label these two forms of differential treatment positive and negative respectively. By *positive differential treatment*, then, I believe we should understand the performance of two actions of different action-types that will disparately affect two persons because of their relevantly similar welfare-constitutive conditions.[22] By *negative differential treatment*, by contrast, we should understand the performance of two actions of similar action-type that will disparately affect two patients because of their relevantly different welfare-constitution. In a catchphrase, *positive differential treatment*, fails because it does not treat discriminatees equally, while *negative differential treatment*, fails because it does not treat discriminatees as equals.[23]

Returning to *Mogul Misogynist*, let us look at the second variation and see if in employing this conceptual distinction we can tell both what is going on and what the difference between the form of discrimination going on there and in the first variation is. First, I believe that much of our intuitive discomfort about the actions of

---

[22] A slightly clumsy term, perhaps, by welfare-constitutive conditions I mean to encapsulate the totality of those physical, psychological and social conditions that have bearing on how their welfare will be affected by the actions of the agent.

[23] The phrasing here borrows, of course, on Ronald Dworkin's famous distinction. (Dworkin, 2000) But the risk of a catchphrase is its potential to mislead. Properly speaking, neither form treats the discriminatees as equals in Dworkin's sense, and neither treats the two equally in the sense I have here suggested.

the Mogul is driven by our concern for the harm likely to be caused to the women involved. This, as I have suggested above, may be because women as a group are socially salient in a way that makes them vulnerable to symbolic harms through displays of differential treatment. Whether wittingly or inadvertently differential treatment of women risks sending a signal that women are in some respects inferior, at once offending women and reinforcing misogynist beliefs and preferences. Furthermore, women may still be, and to the best of my knowledge still are, even in the most progressive countries, socio-economically vulnerable compared to men. Women will, on average, have fewer disposable means, lower income, fewer employment-opportunities, etc. All of which means that a woman will, on average, be harder struck by the loss of her job than a comparable man.[24] In fact, I believe that if we assumed a situation where neither of these conditions were true, then we would, and should, be no more worried about the women of the *Mogul Misogynist* than we are about the less-than-175cm'ers, or indeed whether the mogul's actions disparately affect persons with 3-syllable first names, those who prefer cats to dogs

---

[24] Some might wonder at this point whether one and the same action could at once discriminate against several different groups. This, I think, is both possible and plausible, especially when we allow that negative discrimination does not depend on proxies but on failures to distinguish in appropriate ways, any number of which can occur during a single action. Thus, it is quite possible e.g. that one single action of firing a group of employees is simultaneously discrimination against group A because of their special economic vulnerability, group B because of their special socio-cultural vulnerability and group C because of their special psychological vulnerability, any of which groups can be either partially or completely coextensive.

or those whose birth-year according to CE-reckoning is odd rather than even.

It is worth emphasizing also that in this perspective we can explain why gender is at once, under normal circumstances, morally irrelevant and relevant to cases of discrimination like *Mogul Misogynist*. Properly speaking, the group with which we are concerned is not women, even if they happen to figure disproportionately in the group of those hit hardest by the differential treatment, but those persons, men *or* women, who are harmed by differential treatment. Women, if this is the case, can serve at most as a proxy trait, directing attention to where we might start looking for those who have been disparately affected. What we have then is a case of discrimination against those with particular vulnerabilities, rather than a case of indirect discrimination against women.[25] And further, the wrongness in some cases of disparate impact occurs not because the agent applies a morally neutral distinction in an amoral way (because it disproportionately affects a particular group), but because he fails to appreciate and act upon a morally *relevant* distinction, and in so doing causes avoidable harm.

This explains only how the first and third variations are alike, however. The difference in types of differential treatment involved

---

[25] For a related, but much overlooked, point that uses an analysis along these lines to explain the disparity between racial discrimination and affirmative action, see (Nickel, 1972).

127

explains how they differ. Thus, in the first variation the Mogul engages in positive differential treatment, retaining men and firing women, an action that at once risks causing symbolic harm and reduces the economic prospects of a socio-economically vulnerable group.[26] In the second (and third) variation(s), the mogul engages in negative differential treatment, firing both men and women under 175cm, an action which like in the first variation reduces the economic prospects of a socio-economically vulnerable group and risks causing symbolic harm. There are two significant differences at stake: the first is conceptual, and concerns the difference in the form of differential treatment. Note also how the causal relation between the possession of trait T and the differential treatment changes. In both cases possession of trait T is causally related to the resulting discrimination, but in the case of positive discrimination because it is a necessary part of the explanation of the performance of two sets of relevantly different actions, while in the case of negative discrimination because it is a necessary part of the explanation of the occurrence of two sets of relevantly different effects. The second is moral, in that the risk of causing symbolic harm may be, and likely intuitively feels, somewhat less in the third variation. But the morally significant factor is not the particular ratio between the men and women fired. It is the harmful consequences of acting in one way towards relevantly different discriminatees.

---

[26] Some may wish to hold that the mere risk of causing harm is insufficient for wrongdoing. If so we can assume that the risk is realized, i.e. that the actions of the agent do in fact, or will in fact, cause harm.

Let me return to the question of whether the connection between differentially treated and affected traits could be arbitrary. Consider again the lottery case. It seems to me, on reflection, obvious that there can be cases where subjecting persons to a lottery is morally wrong. Consider two examples from the realm of fiction: The comic book villain "Two-face" has as his distinguishing trait the quirk that he routinely flips a coin to decide whether or not to carry out any particular act of villainy, an idea also employed by Cormac McCarthy for his assassin-character in the novel "No Country for Old Men", who offers (some) victims a fair coin-flip and a 50% chance of being let go alive. Any victim, it seems obvious, might rightfully protest against participating in such a lottery.[27] But in such examples it is the use of a lottery in the first place, rather than any particular outcome, that is rightfully the subject of protest. A victim might very reasonably protest at being subjected to a 50% chance of murder, and continue to protest against having been subjected to it, even after the coin came out her way and she "won" her life because (presumably) she ought not to suffer such risk.

What does this mean for discrimination? It means, I believe, that any random composition of persons will be incapable of protesting that they have been subject to indirect discrimination through the outcome of a lottery. But that many non-randomly composed

---

[27] And, in fact, in a memorably scene the assassin's last victim and wife of the novel's protagonist, does just that. Insisting that responsibility for the outcome rests with the assassin, not the coin.

groups might be able to protest against being subjected to lotteries on the grounds that doing so constitutes a form of direct discrimination.

Some might worry that we are now conflating two types of agency with fundamentally different moral qualities. Specifically, some deontological theories are traditionally loath to attribute the same level of wrongfulness to causing harm by acting in a way which is not a material cause of the consequence, or "allowing", as to causing harm by acting in a way which *is* a material cause of the consequence, i.e. "doing".[28] And it might be imagined that a distinction relying on this schism could be used as an alternative way to support the moral distinction between direct and indirect discrimination here.[29] While there are good reasons to be sceptical about the veracity of the doing-allowing distinction in general, we need not go into that both extensive and seemingly intractable

---

[28] I use the term "material cause" here to capture the common-sense notion of one's actions being part of the causal chain of physical events which produces the effect, upon which the distinction traditionally rests, even though the metaphysics of the question are in fact enormously more complicated and I tend towards the view that even on the level of causal explanation it is so difficult to satisfactorily account for the distinction as to render it indefensible. We need not engage with that particular metaphysical chestnut, however, as I illustrate that the issue is separate from the one at hand.

[29] Cf. e.g. Arneson, who for reasons that appear to be inspired by something like this, argues that: "…disparate impact per se is morally inconsequential. If an act or policy is otherwise morally justifiable, the fact that its consequences favor or disfavor some group of people singled out by some morally arbitrary or neutral classification scheme is not alone a consideration that tends to render the act morally unjustifiable." (Arneson, 2006, p.794, p.794)

debate.[30] A potential objection based on it in this context misses the point that the agent *is* doing something which is the material cause of the harm the discriminatees suffer. For a proper case of an agent merely allowing harm through negative discrimination we would need something akin to:

> *The Misogynist Mogul 4.* Having installed assembly belts and fired the female employees, the Mogul ends up deciding that potential discrimination lawsuits will be too much bother, and sells the factory to an equally wealthy friend. This second mogul does nothing to change the situation at the factory, neither re-employing the discharged women or re-establishing conditions which would make it possible to employ them and/or other women.

As this example hopefully illustrates the doing-allowing distinction cuts across the distinction I have introduced. The second mogul "merely" allows an ongoing form of negative discrimination to continue, whereas the first actively instates the conditions that causes it to occur. Both engage in negative discrimination according to the definition I have introduced. Whether the difference between doing something to cause negative

---

[30] My concerns about the moral relevance of the distinction are largely inspired by the classical criticism developed in (Kagan, 2002, p.83-127). (Howard-Snyder, 2007) provides an excellent overview of the debate.

discrimination or allowing it to continue is morally significant is thus a separate issue.

Conclusion

Anatole France famously quipped that: "…la majestueuse égalité des lois [...] interdit au riche comme au pauvre de coucher sous les ponts, de mendier dans les rues et de voler du pain."[31] (France, 1906 [1894], p.118) The sting of the quip lies, I think, in its exhibition of the callousness of formal equality in the face of grossly disparate conditions. We need to think carefully on what it is that troubles us about such callousness, the better to help ourselves and each other avoid it.

If my analysis of the foregoing pages is right, we are dealing with two forms of discrimination that are conceptually different, but not in the way suggested by the distinction between direct and indirect discrimination. Formalizing the definitions derived above we may summarize as follows.

Positive discrimination against T-persons occurs *iff*:

---

[31] This is slightly inaccurate: in truth Anatole France lets a lead character in the novel "Le lys rouge" state the quip as part of a longer rant on the pitiful state of French society. A fairly literal translation would read: "…the majestic equality of the laws [...] prohibits to rich as to poor the sleeping beneath bridges, begging in the streets and stealing of bread."

1) an agent performs two sets of relevantly different actions towards respectively T-persons and non-T-persons,

2) the actions negatively affect the welfare of T-persons,

3) the difference in actions is suitably explained by T-persons possessing trait T, and

4) in so doing, the agent does moral wrong.

Negative discrimination against T-persons occurs *iff*:

1) an agent performs two sets of relevantly similar actions towards respectively T-persons and non-T-persons,

2) the actions negatively affect the welfare of T-persons,

3) the difference in effect is suitably explained by T-persons possessing trait T, and

4) in so doing, the agent does moral wrong.

As I have also attempted to illustrate in the preceding, this is at first a conceptual rather than a morally significant distinction. I see no reason to believe that it should make a difference to the moral wrongness (or rightness) of the agent's actions whether they are best understood as positive or negative discrimination. However, though it may be no more inherently wrong to differentiate while ignoring a morally relevant trait (the differing welfare-constitutive conditions) than it is to differentiate on the basis of a morally irrelevant trait, there are at the very least strong practical reasons

for agents to take such traits into account, and for us to hold them accountable if they do not. Assuming responsibility for the consequences of our actions in this way is likely the best way to ensure that we do not harm other persons. And because differences in consequence are harder to not only predict but to determine than differences in action, there may be reason to pay particular attention to this part of the moral equation.

Where does this leave the concept of indirect discrimination? If, as I have argued, we need moral wrongness to help us classify standard cases as either direct and indirect discrimination, and if as I have argued indirect discrimination is an alternative but less informative way of understanding the background of the moral wrongness of such situations, then while it remains possible to meaningfully describe cases as direct or indirect discrimination it does not seem particularly helpful. We more accurately understand what is going on by employing the concepts of positive and negative understanding, and lose little that I can see in exchange. Indirect discrimination, then, is a meaningful concept. It just happens to be not a particularly interesting one.

Finally, while not in itself an argument for a harm-based account of the wrongness of discrimination, it does at least indirectly lend credibility to such an account, by illustrating how a plausible explanation of the phenomenon ordinarily conceived as indirect discrimination relies on a concern for harm caused to the discriminatees. It also suggests that we may need to take more account of how our actions affect the welfare of others than we

are sometimes wont. This seems to me an entirely desirable conclusion.

Discrimination and Disrespect[‡]

Introduction: Discrimination, disrespect and the bigoted billionaire

Discrimination is a complicated concept, but as a provisional definition an agent discriminates against someone roughly when:

> 1) she treats those with a particular trait (T-persons) differently than she does those without the trait (¬T-persons),
>
> 2) the treatment is disadvantageous to the T-persons, and
>
> 3) the differential treatment is suitably explained by the T-persons' possessing the trait, e.g. because the trait is what motivates the agent to treat them differently.[1]
>
> (cf. Lippert-Rasmussen, 2006a, p.167-174; also Wasserman, 1998; Halldenius, 2005; Lippert-Rasmussen, 2007a; 2007c; Altman, 2011; Segall, Forthcoming)

---

[1] Some will want to impose further conditions, such as limiting discrimination to only potentially affecting a predefined set of groups. I argue against such a condition in "But Some Groups Are More Equal Than Others". Note also that discrimination will often be divided into direct and indirect discrimination. Although I am sceptical of this division too (cf. "Stealing Bread and Sleeping Beneath Bridges"), it is worth noting that I shall largely be concerned with cases that would tend to be classified under the 'direct'-heading in the present.

Note that this definition says nothing about whether and when discrimination is morally wrong. But now consider the following case:

> *The Bigoted Billionaire.* A multi-billionaire buys a company, and meets with the director. His first order: "I want you to go through the personnel-records and replace everybody who is black, female, over 50, muslim or gay with white, male, below 50, non-muslim heterosexuals." The shocked director protests that this would be illegal, but the billionaire informs him that he does this with every acquisition and that his team of elite-lawyers has found ways of doing it without violating anti-discrimination laws. The director eventually acquiesces, and dozens of workers are fired, their vacant positions filled to the billionaire's specifications.

Irrespective of whether or not the billionaire's lawyers can in fact find ways to dodge legal prohibitions most of us intuitively feel that the scenario describes both a case of discrimination and a moral wrong, that is, that the workers who lose their jobs have been discriminated against in a way which is morally impermissible. But explaining why is not as easy as it may first appear. After all, we do not ordinarily assume that workers have moral claims to their jobs apart from those that are created by the employer and workers as contracting parties and normally covered by labour legislation and individual severance conditions.

Employment, we might say, is fundamentally a voluntary association so that employers are as free to hire and fire workers, as workers are to seek and leave employment. Even should it be the case that this common-sensical assumption is mistaken, a moral claim based on the right to employment may not fully explain the wrong in *The Bigoted Billionaire*, because the intuitions of many will likely point to the fact that the way the workers are fired makes a difference to the scenario. Most of us would feel differently and much less strongly disparaging about an owner who selected the group of workers to be fired e.g. by picking randomly or based on performance. What explains the original intuition and this difference?

The debate on discrimination in the philosophical literature has accelerated over the past ten years and contains a number of different attempts at explaining the wrongfulness of discrimination. (cf. Wasserman, 1998; Radcliffe Richards, 2000; Ezorsky, 2001; Halldenius, 2005; Lippert-Rasmussen, 2006a; Vallentyne, 2006; Arneson, 2006; Edmonds, 2006; Lippert-Rasmussen, 2007a; Heinrichs, 2007; Hellman, 2008; Moreau, 2010; Altman, 2011; Segall, Forthcoming) One of the most prominent explanations of what it is that is potentially morally troubling about discrimination is what I shall call the disrespect-account. This attempts to explain the wrongness of wrongful discrimination in terms of disrespect towards the discriminatee, so that what the billionaire does in my example above is morally wrong because in discriminating in this manner he disrespects the fired workers. In the following I aim to explore different versions of the disrespect-

account, outline both common and individual problems and consider potential solutions. My overall purpose is to evaluate how strong the disrespect-account is and as a result thereof whether we should look to it or alternative explanations of what is wrong, when something is wrong, with discrimination.

I do so by first introducing the concept of respect and distinguishing several importantly different understandings and features of it, drawing on Stephen Darwall's classic definition. Next, I briefly present Larry Alexander's bias variation of the disrespect-account of wrongful discrimination and sketch three basic challenges that the disrespect-account faces, concerning the narrow scope of the account, its counter-intuitive implications and its difficulty in specifying the relevant type of misestimation. Jointly I conclude that these make the bias variation implausible. I then review four variations of the disrespect-account, examining how each changes the argument for what is wrong with wrongful discrimination and evaluating whether they do better than the bias variation. I start with the opacity variation based on recent work by Ian Carter, which adds a constraint of negative respect on the agent, and the valuing variation based on recent work by Joshua Glasgow, which adds a requirement of appraisal respect. I argue that neither avoids the brunt of the basic challenges, and that both introduce new difficulties. I then present two accounts that deviate more radically from the bias variation, by Deborah Hellman and Richard Arneson respectively. The first exchanges recognition for expressive (dis)respect, whereas the second at once broadens the scope of relevant beliefs and imposes conditions on the

background of the estimations of the agent. I argue that although both of these are more successful at avoiding the basic challenges, each of them introduces serious new problems, which makes them ultimately no more persuasive. On the basis of this critical analysis I conclude that the disrespect-account cannot currently be said to satisfactorily explain the wrongness of discrimination.

Exploring respect

A convenient starting point for exploring disrespect-accounts of discrimination is Stephen Darwall's influential 1977 article "Two Kinds of Respect". Darwall usefully distinguishes between *recognition respect* and *appraisal respect*. Recognition respect is defined as: "…a disposition to weigh appropriately in one's deliberations some feature of the thing in question and to act accordingly." (Darwall, 1977, p.38) This, although slightly convoluted, contains the key element of 'appropriately' recognizing and responding to some feature of the object in question. Appraisal respect, on the other hand, is mainly attitudinal in that it consists in the positive appreciation of particular positive qualities: "Unlike recognition respect, one may have appraisal respect for someone without having any particular conception of just what behavior from oneself would be required or made appropriate by that person's having the features meriting such respect. Appraisal respect is the positive appraisal itself." (Darwall, 1977, p.38)

Simplifying and focusing the distinction a little, we might say that an agent shows *deliberative recognition respect iff*:

1) the agent recognizes the relevant feature of the object of respect, and

2) gives this feature the appropriate weight in her deliberations

Similarly, an agent shows *active recognition respect iff*:

1) the agent recognizes the relevant feature of the object of respect, and

2) acts in accordance with what possession of the feature requires

Note that, although Darwall takes both to be part of recognition respect, I believe they are conceptually distinct, and that it is at least theoretically feasible that an agent could meet one but not the other, as well as that there could be a moral requirement to one but not the other.[2] Taken jointly however, we can say that recognition respect simpliciter occurs when the shared first condition and both of the two distinct conditions are met.

Finally, an agent shows *appraisal respect* iff:

---

[2] It might be objected that if we assume moral internalism, then it would be impossible for an agent to take the feature into account in her deliberations in the appropriate way, and not act accordingly. However, I believe that moral internalism is sufficiently controversial that assuming it would severely limit the appeal of the resulting concept. And even if we did, it would still be conceivable for an agent to meet the conditions of active recognition respect while failing the deliberative part.

> 1) the agent recognizes the extraordinary quality of the relevant feature of the object of respect, and
>
> 2) adopts an attitude of positive appreciation for the feature[3]

An important common element between the two is that respect is responsive, or object-generated, in the sense that it is concerned with and determined by features of the thing respected. As Dillon puts it in a like-minded definition: "When we respect something, we heed its call, accord it its due, acknowledge its claim to our attention."[4] (Dillon, 2009)

---

[3] Note that, as we shall see later in the discussion of Glasgow, we can imagine situations in which appraisal respect is required by recognition respect if e.g. we are morally required to adopt an attitude of positive appreciation for some appropriate trait. Note also that, on this definition, it is quite possible to show appraisal respect for the wrong features. E.g. hardened criminals may appreciate and respect the cruelty, cold-bloodedness and disregard for the law of their partners in crime, and although we may condemn such attitudes, and argue that they ought not respect these things, they seem to me clear instances of appraisal respect. It is less clear what to think of cases where the agent mistakenly recognizes something as being of extraordinary quality (i.e. misrecognizes), or recognizes something as being of ordinary quality, but adopts a positive attitude of appreciation nonetheless. However, I leave these complications for devotees of the concept to puzzle over.

[4] Darwall himself stresses this point in his most recent writings, by insisting that his characterization of recognition respect in the 1977 article misses its second-personal character: "This makes respect something one can realize outside of a second-personal relation; one need only adequately register a fact about or feature of someone: that she is a person." (Darwall, 2006, p.131) The point here is that it is not sufficient simply to act in the right way, not even acting the right way while recognizing the existence of the claims one acts in accordance with. It is necessary, for it to be respect on Darwall's account, that one acts for the right

Next, it is worth distinguishing a further feature of Darwall's suggested definition. In his phrasing recognition respect is essentially a question of disposition, in that it involves a specific form of moral reflection which one must be both motivated to engage in and the results of which one must be willing to act upon. In many circumstances, I believe that this is intuitively insufficient, so we can distinguish the dispositional character of Darwall's definition of recognitive respect, which we might call *subjective*, from an *objective* form in which being respectful *just is* "weighing appropriately in one's deliberations some feature of the situation and acting accordingly". The subjective form allows the agent to fail, either in correctly deliberating or acting, so long as she is suitably *disposed* to do so. The objective form sets the bar higher in one respect, although it does not require that one is disposed in the relevant way. However, it seems to me that *pace* Darwall we might want to say that an agent who, despite being disposed to the

---

reasons, reasons which are generated by one's recognition of the relation one stands in to the person who raises these claims and her ability to raise them qua her dignity. (Darwall, 2006, p.140-141) Otherwise, one might be said to respect certain features of the situation, or even certain moral claims generated by a person, but not to respect the person herself as a person: "To respect someone as a person is not just to regulate one's conduct by the fact that one is accountable to him, or even just to acknowledge the truth of this fact to him; it is also to make oneself or be accountable to him, and this is impossible outside of a second-personal relation." (Darwall, 2006, p.142) In other terms, it is essential to the concept, for Darwall, that respect generates agent-relative duties. We need not take a stand on this issue however, as the problems that emerge in the discussion to follow do not rely on the relevant duties being agent-neutral. Darwall's development of a second-personal theory changes, I believe, the normative basis rather than the conceptual content of respect.

contrary, weighs a relevant feature in the appropriate way and acts accordingly is in some sense respectful.

One final distinction concerns what is required for an agent to be disrespectful. Consider that, on any of the versions of recognition respect distinguished above it is not immediately apparent whether recognition *disrespect* is best understood as a failure to properly take account of morally relevant features or the wrongful taking into account of morally irrelevant features (or, having the requisite dispositions on the subjective version). On the first interpretation, the deliberative element of objective recognition disrespect consists in that:

> 1) the agent fails to recognize, or
> 2) gives inappropriate weight in her deliberations
> to some morally relevant feature of the patient[5]

Call this *positive* recognition disrespect, as it violates the requirement that we recognize and weigh the relevant feature(s). On the second interpretation, the deliberative element of objective recognition disrespect consists in that:

---

[5] Though cognitively these might work differently, principally they are simply the difference between mistakenly ascribing a weight of zero, and ascribing a mistaken but greater than zero weight to the feature. This becomes even clearer once we consider negative disrespect below, which is mistakenly ascribing a greater than zero weight. Note also that I am not here restricting misestimations to being either under- or overestimations, a point to which I shall return later.

1) the agent mistakenly recognizes the object as possessing a morally relevant feature it does not possess, or

2) gives inappropriate weight in her deliberations to some morally irrelevant feature of the patient[6]

Call this *negative* recognition disrespect, as it violates the requirement that we refrain from recognizing and giving undue weight to nonexistent or irrelevant features.[7]

Recognizing what and why?

In the above, I have discussed respect broadly enough that it can essentially pertain to any feature of any object, from the aesthetic quality of works of art to the justice of legal institutions. Proponents of a respect-oriented ethics will, however, typically focus on a particular sense of respect peculiar to morality: respect for moral patients, which in Darwall's Kantian perspective happens to be persons: "Persons can be the object of recognition

---

[6] The object of respect need not be a moral patient in the case of negative disrespect. A moral agent who mistakenly believes that rocks have the same moral worth as persons is (presumably) grossly overestimating the moral worth of rocks; rocks are not moral patients, i.e. they have no moral worth, so that ascribing them *any* moral worth is a failure to appropriately recognize and weigh their worth.

[7] On consideration, the most plausible understanding of recognition respect seems to me to be that to comprehensively respect some thing one would have to do neither, which implies that the deliberative element of respect consists in recognizing and appropriately weighing *all and only* those features of the object that are morally relevant. However, we shall explore variations based on the different understandings below.

145

respect. Indeed, it is just this sort of respect which is said to be owed to all persons. To say that persons as such are entitled to respect is to say that they are entitled to have other persons take seriously and weigh appropriately the fact that they are persons in deliberating about what to do. Such respect is recognition respect; but what it requires as appropriate is not a matter of general agreement, for this is just the question of what our moral obligations or duties to other persons consist in. The crucial point is that to conceive of all persons as entitled to respect is to have some conception of what sort of consideration the fact of being a person requires." (Darwall, 1977, p.38) It is worth noting the specific features of the quality under consideration here. What Darwall is concerned with is what we might call the patient's 'moral worth', and concerns just whether and if so to what extent the object under consideration is in fact a moral patient. But this can be conceived of dualistically or as a concept that allows of degrees. Thus, arguably, on some consequentialist accounts different animals will have different moral worth relative e.g. to their level of consciousness and capacity to experience pain and pleasure. On the other hand, most Kantian accounts, tracing moral worth to notions of dignity or autonomy, will tend to take a dualistic approach and consider moral worth a concept that does not allow for degrees; one either is, or is not, a moral patient, depending e.g. on whether or not one is an autonomous being. Compare this type of feature to other morally relevant features such as desires, capabilities, needs, rights and deservingness, which are features *of* moral patients. It is the strength of the claims these

146

features stake on moral agents that is potentially discounted according to the patient's moral worth.[8]

The next thing to notice about this understanding of respect is that, although the concept of respect has obviously had a more prominent role in a broadly Kantian tradition of moral philosophy there seems to be no obvious reason why consequentialist ethics could not involve respect described in the same terms. Because the definition of respect strictly speaking does not rely on a specific understanding of the relevant object(s) of respect and leaves what the appropriate weighing and resulting actions are unspecified, it is consistent with the definition to say e.g. that a utilitarian shows recognition respect of a conscious being when she applies Mill's famous dictum, apocryphically attributed to Bentham: "everybody to count for one, nobody for more than one" (Mill, 2007 [1871], p.105), that is, just when she counts the impact of her actions on that being's wellbeing as an equal part of the total wellbeing which she seeks to maximise. To do so would seem to meet the criterion

---

[8] Some Kantians might want to hold that *only* persons are worthy of respect, i.e. that although an agent *can* respect many different things, the moral duty to respect extends to all and only persons. This makes sense when applied to those features of moral patients that ground substantive duties, such as rights or interests. It seems strange, however, when applied to moral worth, because recognizing someone as a moral patient presupposes deliberating about the wider group of potential moral patients. Bluntly put, we cannot start respecting a person as a person until we have decided whether or not what we are faced with is in fact a person. Kantians will need to decide, it seems to me, whether to maintain that there is no moral duty to recognize persons as such, but only to respect those recognized as persons, or that there is both a moral duty to recognize persons as such, which implies recognizing non-persons' (lack of) moral worth, and to respect those recognized as persons.

of "a disposition to weigh appropriately in one's deliberations some feature of the thing in question and to act accordingly", given that the relevant features on a utilitarian understanding is the being's capacity to experience pleasure and pain, the correct weighing is impartial and the appropriate action is that of the available alternatives which maximises the total good.[9]

It is on a different level, therefore, that the distinctive element in Kantian understandings of respect emerges. On standard accounts of consequentialism, the duty to treat others respectfully in the sense here defined is at most derivative.[10] As Phillip Pettit has argued it is not generally speaking true that consequentialism requires the agent to *think* as a consequentialist, e.g. by weighing all persons' wellbeing equally and attempting to select for total maximisation, it is simply true that all persons *do in fact* count equally in just this way, and that the right action *is in fact* that which maximises.[11] Thus: "…while it is appropriate to assess or evaluate an option by reference to the values of its prospects, it may not be appropriate for an agent to use such assessment in his deliberation. It may be better for him – it may improve his chances of getting a desirable prognosis, for example – if he restricts his deliberation,

---

[9] I set aside in this broad use of 'utilitarian' important variations on this definition, including notably subjective- and rule-utilitarianism. The same point applies, however, to such variations, with the requisite alterations.

[10] There will, presumably, be versions of both subjective- and rule-utilitarianism where this is not true, but as above I set these aside.

[11] This is, I admit, putting it too strongly. Consequentialism might employ other principles for weighing outcomes than impartial maximising. But the same point stands for forms of consequentialism that employ e.g. prioritarian weighing.

making his decisions by using certain rules of thumb or whatever."[12] (Pettit, 1989, p.119) Whether or not agents ought to think about their actions in consequentialist terms will itself be decided by whether this is the procedure that will lead such agents to achieve the best consequences.[13]

We can distinguish therefore between the claim that recognition respect, i.e. (adopting a disposition to) weighing appropriately in one's deliberations some feature of the thing in question and to act accordingly, is a contingent and instrumental duty, as it may be for consequentialists in some even if not in all realistic scenarios, and the claim that it is an intrinsic duty.[14] Darwall, and presumably

---

[12] Similarly, Derek Parfit has famously argued that consequentialism is, in his formulation, 'indirectly self-defeating' because there will be situations in which coordination-problems will make it preferable for consequentialists, as measured by the requirements of consequentialism itself, to adopt non-consequentialist norms, i.e. to be disrespectful on the terms of the present account. (Parfit, 1984, p.24-67)

[13] This does not mean that consequentialists are concerned with what I have called objective recognition respect. Even that requires that the agent is responsive to and deliberates about the morally relevant features in the appropriate way, in addition to acting in accordance with such deliberations. If we assume utilitarianism, doing so will necessarily lead an agent to act in the way that is morally required by utilitarianism, but this is morally required for independent reasons, and the agent's deliberations and responsiveness do not contribute anything to the moral status of the situation. In short: there is no requirement to be responsive and deliberate in a specific way, only to act in a specific way, even if that way is necessarily coextensive with the actions that *would* be the result of perfect deliberation.

[14] An objection might hold that in those situations where consequentialism requires the agent to not deliberate along consequentialist lines, rather than calling this disrespect we should understand it as a form of 2nd-order respect, in which properly taking account of the features of the situation, that is, being respectful, requires acting in a way that will not in the same way properly take

many Kantians, will prefer to hold *pace* consequentialists that the latter is true, and that we have a duty to be respectful even e.g. when doing so perversely does not promote the value(s) which respect is appropriately directed towards taking into account. In Phillip Pettit's analysis, this form of non-consequentialism denies that we can assess the rightness of choosing an option simply in terms of its promoting a given value, holding rather that we are morally required to be respectful, irrespective of what outcome this produces. (Pettit, 1989, p.118-121; cf. also Brown, 2011, for a concise recent statement of this difference) However, evaluating the soundness of this claim will take us into intractable differences between the consequentialist and deontologist camps, and I shall not pursue it here. Let us rather assume for the sake of argument that there could be an intrinsic duty to recognitive respect, and see whether or not this helps us explain the wrongness of discrimination.

Discrimination and disrespect

Consider first, what it means for discrimination to be wrong. Clearly, it is not sufficient that wrongs are involved in an act of discrimination. If a psychopath kills one person and spares

---

account of the features of the situation, that is, being disrespectful (on the terms initially defined). This would describe the difference as being between consequentialists who would subsume the principle of respect under its own standards, that is, measure when to be 1st-order respectful by the identical but higher level standards of 2nd-order respect, and non-consequentialists who would not. I consider this a compatible description of the same phenomenon, and so leave it undecided whether one is in some way preferable to the other.

another, picking her target because the first is a T-person while the second is a ¬T-person then, although she discriminates against the victim by treating her differently, the discrimination is a different and much less serious wrong, if it constitutes a wrong at all, than the isolated act of killing an innocent.[15] The wrongness of discrimination is therefore not explainable by reference to the wrongness of her killing – this wrong is *extrinsic* to the wrong of discrimination.[16] What we are looking for is rather an *intrinsic* wrong, that is, one either caused *by* – not *while* – discriminating or one inherent in (some forms of) discrimination itself. Thus, if we imagine subtracting the wrongness of the killing from the situation described above, perhaps by comparing it with the wrongness of non-discriminatorily killing an innocent, whatever wrongness remained might be attributable to discrimination. (cf. Hellman, 2008, p.15-16)

---

[15] It might seem to make a difference here whether she has already decided to kill one person, and simply uses her (dis)preference for killing T-persons to select the victim, or whether her (dis)preference for T-persons is what motivates her to kill, so that barring the person possessing the trait, nobody would have gotten killed. But this is a mistake. It might be said that the animosity she feels against T-persons which motivates her to kill also motivates her to discriminate, but this does not make the wrong of her killing part of any wrong committed by discriminating.

[16] The same point was illustrated above in *the bigoted billionaire* when I claimed that the wrongness of the billionaire's actions is intuitively not reducible to any wrong committed by discharging employees. Even if this is independently wrong it seems that there is a relevant difference between situations where this wrong is committed and discrimination occurs, and those where the wrong is committed and no discrimination occurs.

Suppose now that there is a feature of some cases of discrimination, which is presumptively morally wrong and which is present in all and only cases of morally wrongful discrimination.[17] Could this feature be said to explain the wrongness of discrimination? It might, if it also met two further conditions: first it would need to be plausibly related to discrimination, in such a way that we can label it an intrinsic wrong, that is, its coextensiveness with wrongful discrimination must be determinably non-accidental. To the extent that it is not demonstrably so related, this weakens the plausibility of an account which relies on it. And second, it would need to exclude and better explain the wrong than features emphasized by alternative accounts. If it excluded and better explained than other accounts, then we could rule out those alternatives and consider this feature to fully explain the wrongness. Alternatively, it might only supplement other accounts, and so not supply the whole explanation, but at least partly explaining what is wrong with discrimination. This is where disrespect can enter the picture: if disrespect is a feature common to all cases of wrongful discrimination, and if disrespect is presumptively morally wrong, then we might have at the least a reasonable assumption that disrespect is what is morally wrong with discrimination.

---

[17] That is, it is never present in morally permissible discrimination, although we do not need to rule out its presence in both permissible and wrongful non-discrimination.

The bias variation: disrespect from misestimation of moral worth

An initial attempt at linking discrimination to disrespect might point to the fact that one of the central features of wrongful discrimination, well illustrated by classic cases of racism, misogyny, homophobia, etc., is that the discriminator perceives the discriminatee to have a lower moral worth than other persons. Such mistaken judgements that the discriminatee has a lower moral worth, or 'bias' in Larry Alexander's phrasing, fail to appropriately weigh a central moral feature of the discriminatee and thus to recognitively respect her on the terms given above. In Alexander's words: "[Nazi biases against Jews] were intrinsically morally wrong because Jews are clearly not of lesser moral worth than Aryans. When a person is judged incorrectly to be of lesser moral worth and is treated accordingly, that treatment is morally wrong regardless of the gravity of its effects. It represents a failure to show the moral respect due the recipient, a failure which is by itself sufficient to be judged immoral." (Alexander, 1992, p.159)

This suggests the following definition in line with the terms that I have applied so far: An agent engages in *biased (positive objective recognition) disrespect* discrimination against T-persons *iff*:

> 1) the agent treats T-persons differently than
> ¬T-persons,
> 2) the differential treatment is disadvantageous
> to T-persons, and

3) the differential treatment is caused by her failure to recognize and appropriately weigh the moral worth of T-persons[18]

As Richard Arneson observes, an implication of this type of mental-state account is that one can wrongfully discriminate while doing what would otherwise be permissible or even morally required, if one does so for reasons that are inappropriate to the situation at hand *and* instances of the requisite flaws: "I submit one can be guilty of wrongful discrimination when one treats a person morally appropriately (so far as one's behavior is concerned) and better than one would have done had one not been moved by negative attitudes or bias against the group of which one holds the person to be a member. Consider this example: One treats a person better than one otherwise would have done from animus or prejudice against persons of that type. One says to oneself, "I'd better pay what I owe to Sally, because she is a pushy Jew, or an uppity black, or whatever, and would respond more aggressively to not being paid than other persons of better type.""[19] (Arneson, 2006, p.779, note 14)

---

[18] Note that condition 2 is added to Alexander's account. I believe, however, that it is necessary for us to be capable of speaking about 'discrimination against'. This does not imply, however, that 'discrimination for' T-persons (where the differential treatment advantages them) or discrimination between T-persons and ¬T-persons (where the treatment favous nobody) is morally different from discrimination against. Presumably, on a disrespect-account such cases will be wrong for exactly the same reasons.

[19] I return to explore Arneson's slightly more complicated version of the disrespect-account in the culpability variation below.

This will likely be considered a strength of the view by its proponents, as it makes room for labelling as morally impermissible actions that it would otherwise seem difficult to prohibit on deontological grounds. Thus, many cases of discrimination operate in the space of options where, generally speaking, agents are free to pursue whatever choices they desire because no moral rights bear on the issue: "Consider whimsical hiring. I am hiring persons to work in a doughnut shop I own. There are several other doughnut shops in the neighborhood, so it will not be a great loss to any actual or potential customer if my doughnut shop is not run as well as it might be. I announce that I will respond to the applications according to my subjective mood and select an applicant to be hired by arbitrary whim. This does not seem to be in the ballpark of wrongful discrimination." (Arneson, 2006, p.784) Yet, and this is the point, if we change the scenario to something resembling *the bigoted billionaire*, where the agent has an intention to not hire e.g. blacks and latinos, the inclusion of this morally corrupt intention would seem for many to make the action itself discriminatory, in the pejorative sense of being morally impermissible.[20]

---

[20] Arneson considers and dismisses the possibility that we should revise our intuitions about whimsical hiring not being discriminatory. I concur with Arneson on this point; it seems to me that on any reasonable deontological account there must be some room for options of differential treatment for morally irrelevant reasons. So long as this is the case, the strong meritocratic norm that would invalidate Arneson's example does not obtain. (Arneson, 2006, p.783-785) Things look different, of course, if one assumes a consequentialist perspective. But holding this against Arneson and other proponents of the

The account meets the conditions I sketched above, in that it identifies a wrong-making feature which is non-extrinsic and non-arbitrarily connected to discrimination. And, it is capable of at least potentially accounting for cases like *Bigoted Billionaire*, in so far as the firing of the employees is based on the kind of misestimation of moral worth which constitutes disrespect. It looks to be a promising candidate for an account of what is wrong, when something is wrong with discrimination.

The basic challenges: (mis)estimation, contrary intuitions and the speciesist scientist

As I have shown above, the failure to show recognition respect, particularly what I have labelled the bias-variation, is an apparently plausible explanation for what is wrong with discrimination. However, the variation faces a number of serious problems that I shall outline in the following. The first concerns the character of (mis)estimation, where the account has problems non-arbitrarily specifying the relevant kind. The second follows from a powerful counter-example developed by Kasper Lippert-Rasmussen. Finally, the third contests that disrespect cannot plausibly explain large groups of cases that intuitively seem to involve wrongful discrimination.

---

disrespect-account here is simply to deny the deontological foundations of the account.

As concerns the first, recall that the third condition of this version of the disrespect-account requires that:

> 3) the differential treatment is caused by her failure to recognize and appropriately weigh the moral worth of T-persons

Now, Alexander initially suggested that wrongful discrimination was based on and explained by the undervaluation of the discriminatee's moral worth. As Lippert-Rasmussen puts it: "On this account, the problem with discrimination is not that the person who discriminates is too respectful of some (those in favour of whom he discriminates) but rather that this person is disrespectful of others (those against whom he discriminates)." (Lippert-Rasmussen, 2006a, p.180) This does not look promising. If the disrespect-account wants to hold that disrespect is the underestimation, rather than the misestimation of moral worth, then it is forced to implausibly claim that there is a morally relevant difference between treating person A worse than person B because one underestimates the moral worth of A, and treating person B better than person A because one overestimates the moral worth of B, even though both seem to be valid descriptions of similar cases of potentially wrongful discrimination.[21] Thus, a

---

[21] Consider a situation where, when confronted with the claim that his firing of the employees is wrongful discrimination, the Bigoted Billionaire admits to having misestimated the moral worth of somebody. "However", he says, "I did not *underestimate* the moral worth of those I fired; I *overestimated* the moral worth of those I replaced them with. My actions were based on the sadly mistaken

157

better alternative is to hold that: "…discrimination is bad because it involves a *false* representation of someone's moral status, where the falsehood need not consist in representing someone as having a lower moral status than he in fact has and where that person need not be the discriminatee, i.e. the immediate object of discrimination." (Lippert-Rasmussen, 2006a, p.181, my emphasis; cf. also Feinberg, 1974, p.306) This approach is consistent with different formulations of the principle in Alexander's article, and how I have summed up the idea in the above definition. Wrongful discrimination, Alexander later claims, is mainly wrong since: "…biases premised on the belief that some types of people are morally worthier than others are intrinsically morally wrong because they reflect *incorrect* moral judgments."[22] (Alexander, 1992, p.161, my emphasis)

---

belief that white, middle-aged heterosexual men had a higher moral worth than they in fact have, whereas my perception of the moral worth of everyone else was and is accurate. I now realize that everyone in fact has the exact same moral worth, but since my actions were not based on underestimating the moral worth of anybody, they were not wrongfully discriminatory, and as such require no rectification." Most of us would, I believe, find this type of defence both unpersuasive and bizarre.

[22] The ambiguity is not resolved, however. It persists when Alexander concludes that: "Biases - except for those reflecting close personal ties that are so central to one's identity they amount to "biases" in favor of one-self are paradigmatically intrinsically immoral. Biases rest on erroneous judgments of others' *inferior moral worth*, and they insult and produce justifiable resentment in the dispreferred, harms that are unnecessary additions to the inevitable harms of being dispreferred." (Alexander, 1992, p.192, my emphasis) Note in this context that the harms here introduced ought on Alexander's account to be considered relevant only to the *extrinsic* wrongness of the act. In addition, although he concedes that there are cases where discrimination is wrong by virtue of the

Consider, however, the character of the misestimation necessary for disrespect to be morally wrong. Such misestimation might be mistaken absolutely or comparatively, that is, either by misestimating the patient's moral worth in relation to how much worth the patient herself actually has, or by misestimating it in relation to how much worth other patients have. Which of these is required to be disrespectful? Consider first the possibility of absolute misestimation. Lippert-Rasmussen argues that this does not work, because it is the comparative mistake which plausibly explains the wrongness in cases of discrimination involving both: "An individual, X, may incorrectly consider himself morally more worthy than Y and at the same time consider himself *and* Y to be morally more worthy than they in fact are. Presumably, X's bias reflects incorrect moral judgement, but what makes the bias disrespectful is X's incorrect judgement about relativities, not his incorrect judgement of absolute moral worth." (Lippert-

harm it causes, Alexander also posits a high threshold for the wrongfulness of this harm, which implies that the intrinsic wrong of disrespect forms the core of what is wrong when something is wrong with discrimination. Cf. Alexander: "Discriminatory preferences are extrinsically morally wrong if their social costs are large relative to the costs of eliminating or frustrating them." (Alexander, 1992, p.219, p.219) Arneson, who we shall turn to momentarily, offers a similar qualification: "Of course, acts that are not intrinsically morally wrong may become morally wrong for extrinsic reasons. This is so when an act takes place in circumstances where it causes bad consequences to an extent that outweighs its intrinsic innocence." (Arneson, 2006, p.790-791)

Rasmussen, 2006a, p.181) Comparative misestimation may seem the better option then.[23]

But there are still situations, it seems to me, where proponents of a disrespect-account would want to claim that absolute misestimation constitutes a moral wrong. Imagine a situation in which the agent considers herself and everybody else to be equal, but vastly less morally worthy than they in fact are. Obviously, given the equal assessment such a misestimate cannot give rise to discrimination, but it might still be morally wrong. And indeed it seems to me that any plausible account of disrespect would have to hold that such a person – someone who considers everybody to be morally less worthy than they are, to the point, perhaps, of considering nobody to have any moral worth at all – is the paradigmatic exemplar of a disrespectful person.[24] This spells

---

[23] Note that this distinction is different from, and to some extent cuts across the distinction between positive and negative disrespect I introduced above. We shall return to an attempt to base the wrongness of discrimination on negative disrespect in the section on the opacity variation below. Those familiar with Joel Feinberg's work on comparative justice may also want to note that on his account judgements of moral worth, whether comparative or non-comparative, fall within the scope of *non*-comparative justice, although discrimination involves the requisite second comparison to bring the issue under comparative justice. (cf. Feinberg, 1974, p.302-304)

[24] Note that this does not entail that the duty of respect is itself comparative. In both recognitive and appraisive cases it amounts to responding in the correct way to traits that pertain to an individual, independently of how one responds to other individuals. It is possible of course, e.g. on egalitarian normative theories, that the relative position of an individual is one of the moral features that must be taken into account, just as it is possible that appraisive respect is properly directed towards e.g. the relatively superior skill, virtue or intelligence of an individual, but this makes the *object* of respect comparative position, rather than making respect itself comparative. To be properly comparative, we should have

trouble, however, because we can conceive of cases where absolute misestimation is what gives rise to discrimination. Suppose for instance that we are distributing a cost among two patients, A and B, with unequal moral worth – e.g. we can save only the dog or the owner from the burning building – and we are mistaken about the moral worth of both in absolute but not in relative terms. That is, while we misestimate the moral worth of both, we do so in a way which leaves our estimate of their comparative moral worth correct.[25] Suppose further that it is true that had we not absolutely misestimated the moral worth of A, who has the comparatively lower worth, we would have mistakenly believed A to have the comparatively higher (or perhaps, merely equal) moral worth.

---

to say that respect does not consist in (a disposition to) weighing *appropriately* the relevant features of the situation and act accordingly, but only to weigh *equally*, or in the same way, the relevant features of the situation and act accordingly. But this understanding implies that a person who is strongly non-comparatively disrespectful can be comparatively respectful, e.g. because she always takes *none* of the relevant features into consideration. This strikes me as highly implausible.
[25] Some might object that moral worth is, in at least some Kantian theories, a dualistic concept which would not allow for this type of combination of absolute and comparative misestimation. One either has, or does not have, moral worth, but it makes no sense to say that a comparative estimate could be correct if the absolute estimate is not and the two beings compared have unequal moral worth, because one of them will in fact have zero moral worth, and no absolute misestimate will preserve the comparative ratio. In the example, this corresponds to the dog having zero moral worth, and the owner one. Clearly, no value other than zero will preserve this ratio (or, strictly speaking, any value other than zero will introduce a ratio that is not there). Admittedly, if this was true, it would diminish the problem somewhat, but although I confess to finding the notion of dualistic moral worth extraordinarily implausible I shall not argue against it here. Even granting dualistic moral worth, the challenge remains to account for the discrepancy between our intuitions that absolute misestimation is the basis of disrespect in some cases and comparative misestimation in cases of discrimination.

Hence, the differential treatment of A is counterfactually caused by our absolute misestimation of A's moral worth. And if, as I suggested above, disrespect accounts must hold absolute misestimation to constitute a wrong in itself, then we now have case of wrongful discrimination based on absolute misestimation.[26] But, as I argued earlier, if neither underestimation, comparative misestimation or absolute misestimation can cover all plausible cases of disrespectful discrimination, disrespect fails to plausibly explain the moral wrongness of discrimination.

Furthermore, as Kasper Lippert-Rasmussen has argued, this version of the disrespect-account faces a serious problem in that it seems to turn our standard intuitions about the way badness is tied to belief on its head. Normally it would seem to most of us that if the beliefs of an agent influence the permissibility of the action at all, an act must be less wrong if the agent does it while holding a mistaken belief which implies that what she is doing is in fact right.[27] Thus, causing harm to someone because one has failed to

[26] Note that the plausible counter-argument, that this cannot constitute wrongful discrimination, since the agent actually discriminates in the way she *ought* to discriminate is not available to the proponent of the disrespect-account, since she is wedded to the notion that the wrongness is explained by the background of the action, in the shape of the agent's estimations, rather than by the action itself.

[27] Assuming that beliefs affect permissibility, my pouring arsenic into your coffee because I mistakenly believe it to be sugar surely makes the action *less* morally wrong than it would be had I known it to be poison. The alternative here is, of course, to say that the beliefs, attitudes etc. of the agent do not affect the wrongness of the action, but the blameworthiness of the agent. But this is not consistent with the recognitive disrespect-account.

properly recognize their moral worth, in that one holds the mistaken belief that the patient one is harming has no (or less) moral worth and that causing it harm is therefore morally permissible, seems, if anything, less bad than inflicting similar harm while recognizing that the subject has equal moral worth and holding the accurate belief that one's action is therefore morally wrong.[28] Consider this slightly simplified version of an example given by Lippert-Rasmussen, which we might call:

> *The Speciesist Scientist*: A pharmaceutical company discriminates against animals by inflicting horrible pain on hundreds or even thousands of them to provide a very small benefit to a small group of humans, e.g. the ability to buy and wear a new perfume without suffering a small risk of having a mild allergic reaction.
>
> *Researcher 1* misestimates the moral worth of animals, and therefore falsely believes that her actions are morally permissible.
>
> *Researcher 2* correctly estimates the moral worth of animals, and therefore correctly believes that

---

[28] Note that, as Lippert-Rasmussen makes clear, we must set aside as extraneous factors both the possibility that the first action will cause *greater* harm, because it may be insulting to be harmed in this way, the situations in which the agent is not epistemically justified in holding the mistaken belief and the potential culpability that would result if the reason the agent holds mistaken beliefs is that she manifests certain epistemic flaws that are independently worthy of condemnation. (Lippert-Rasmussen, 2006a, p.182-183)

her actions are morally impermissible. She pursues them regardless.

As Lippert-Rasmussen concludes, this poses a serious problem for the disrespect-account of discrimination, because: "…if one's discriminatory activities are less bad when they are accompanied by an underestimate of the moral status of the discriminatee than they are when they are accompanied by a correct estimate of the discriminatee's status, it follows that discrimination cannot be bad simply because it reflects an incorrect judgement of moral status." (Lippert-Rasmussen, 2006a, p.183)

Finally, it is worth noting that the disrespect-account does not seem capable of explaining a large number of plausible cases of non-biased wrongful discrimination. It is perfectly possible, for example, to imagine an agent who at one and the same time genuinely believes:

> a) that Blacks and whites are equally morally worthy, and
>
> b) that Blacks are given to e.g. laziness, sexual aggression and stupidity

Supposing that *the bigoted billionaire* was such an individual, his discrimination against Blacks (and the other groups, with the requisite variations) would be both rational and recognitively respectful, but not, we would presumably want to say, morally permissible. Partly, this may be because nobody today is epistemically justified in believing b), so that holding this belief

might be said to be disrespectful with respect to a different feature, because it responds inappropriately to evidence regarding the racial traits of Blacks, but this does not imply a perception of unequal moral worth or a failure to recognize Blacks as persons.[29] Someone believing b) could theoretically, if perhaps only with some difficulty in practice, believe that this made no difference to the respect, appreciation and moral duties she owed blacks *as persons*.[30] In summation, misestimation of moral worth in particular is too narrow an account to be capable of including large numbers of cases that intuitively strike us as obvious examples of wrongful discrimination.[31]

These then are the basic challenges for the disrespect-account. It does not, on the terms I suggested as requirements for a wrong-making feature above, seem a plausible interpretation of what is

---

[29] A possibility pursued by Arneson, cf. below.

[30] If this still seems implausible, consider the analogous situation of children. Presumably, most people (accurately) believe that children, say between the age of five and ten, are stupid, irresponsible and self-centered, at least when compared with adults and even though some might prefer to put the point more delicately. Few would hold, however, that this in any way alters the moral worth of these children. The difference between these two cases concerns the veracity of the beliefs in question, not their compatibility, logically or psychologically, with affirming the equality of moral worth.

[31] Some might counter that a certain level of discrepancy between intuitive cases and a moral theory is both permissible and expectable. In fact, one of the aims of moral theory is to help us revise our intuitive understanding; moral theory is corrective, not simply conservative. I essentially concur, but my claim here is the somewhat stronger that there will be large groups of intuitive cases that fall outside the scope of the disrespect-account. Extensive discrepancies between our considered intuitions and the claims of a theory should be held against the theory, I think, more so the greater the discrepancy.

wrong with discrimination. To solve the problems, an alternative account must at once avoid the ambiguity surrounding misestimation, specify the type of disrespect involved in discrimination in a way that better captures prominent examples and avoid running counter to our intuitions about the relation between mistaken belief and permissibility.

Four variations to meet the challenges

Having laid out a framework for assessing the disrespect-account of discrimination, we can now proceed to look at four variations. As we have seen, Larry Alexander's 1992 article introduced the notion in the shape of the idea that the crucial factor which determines the permissibility of discrimination is whether or not the differential treatment is biased, which on Alexander's account means that it is based on a misestimation of the discriminatee's moral status (Alexander, 1992). Two recent articles have suggested minor variations on the understanding of respect pertinent to the discussion. Ian Carter argues for opacity respect, which includes a threshold above which differences among agents should not be taken into account (Carter, 2011), while Joshua Glasgow argues that the bias variation must be supplemented with a condition requiring appraisal respect (Glasgow, 2009). The disrespect-account is further developed in Richard Arneson's "What is Wrongful Discrimination?", which remains probably the strongest and most detailed version of the disrespect-account (Arneson, 2006). Deborah Hellman's "When Is Discrimination Wrong?" focuses on an expressive interpretation of the argument for disrespect as a wrong (Hellman, 2008). In what follows I shall

review each of these variations, starting with those based on Carter and Glasgow and ending with Hellman and Arneson, before summing up and concluding.

The opacity variation: disrespect from taking the wrong features into account

I have touched upon the possibility of casting the constraint against disrespect as a violation of a negative duty above. Recent work by Ian Carter pursues a conceptualisation of respect along these lines, (Carter, 2011) and it may be worth briefly considering whether redefining respect this way would help proponents of a disrespect-account of discrimination. Carter's general aim is to provide a satisfactory account of the basis of equality, that is, to explain what it is that all moral patients share, and share equally, which makes them subjects of equal respect. In doing so he argues in favour of understanding moral worth as a range property, and for an upper threshold beyond which we are morally obligated to disregard differences in the traits that ground moral worth: "[R]espect for persons consists not simply in the recognition of (variable) empirical agential capacities in certain beings, but in (i) the recognition of their possession of an absolute minimum of those empirical capacities plus (ii) the adoption of the external perspective that is appropriate in the case of any being that is seen as having at least that minimum." (Carter, 2011, p.553)

The type of barrier to deliberation at stake here, is therefore one where the agent *does* take into account a certain feature of the object of respect, but where she limits her assessment of it in a

certain way. Specifically, she makes an estimate of to what extent the object possesses the feature, compares this to a threshold, and for all values above the threshold she then *pretends* that the value is equal to the threshold.[32] Failure to meet this form of respect is not misestimation or taking the wrong features into account, but failure to pretend that a certain range of values are not less than they in fact are.

Admittedly, Carter takes this to be a constraint on practices of equalizing, which tend to take the shape of discriminating *for*. But if we assume that it can also apply to practices that do the opposite, then it seems we have a potential wrong-making feature for acts of discrimination.[33] Carter suggests that when evaluating practices we should then apply: "*The opacity test*: a practice passes

---

[32] Thus, Carter writes: "Let us say that to respect persons in the above way is to treat them as opaque. More precisely, it is to treat them as opaque up to a point, on the scale or scales measuring their agential capacities. Treating persons as wholly opaque, in the sense of completely ignoring their agential capacities, would be too strong, for it would preclude those assessments that are necessary in order to have the reasonable belief that they have any agential capacities at all." (Carter, 2011, p.552) And: "Respect, on this alternative interpretation, is a substantive moral attitude that involves abstaining from looking behind the exteriors people present to us as moral agents. More precisely, while we may see behind these exteriors (for to do so is often unavoidable), if and when we do perceive people's varying agential capacities we refuse to let such perceptions count as among the reasons motivating our treatment of those people. In other words, we avoid evaluating people's agential capacities as an aid to deliberation about alternative courses of action." (Carter, 2011, p.551)

[33] Indeed, I am skeptical that any argument can be made as to the moral significance of the difference between discriminating for and discriminating against. They seem to me to be interchangeable ways of describing the same situation, given that any discrimination for T-persons will be discrimination against ¬T-persons, and vice versa.

the opacity test if and only if the carrying out of that practice neither constitutes nor presupposes any violation of the requirement of opacity respect." (Carter, 2011, p.561) If we further venture that a similar condition could be applied to suitable cases of individual action, it seems we have the basis of an alternative disrespect-account of wrongful discrimination. On such an account an agent engages in *opacity (negative objective recognition) disrespectful* discrimination against T-persons *iff*:

> 1) the agent treats T-persons differently than ¬T-persons,
> 2) the differential treatment is disadvantageous to T-persons, and
> 3) the differential treatment is caused by a) her failure to recognize and appropriately weigh the moral worth of T-persons, or b) her taking into account differences in the traits constitutive of moral worth above the threshold

Interesting as the opacity-account may be in relation to respect, it seems however that opacity disrespectful discrimination fares, if anything, worse than the bias variation. Note first that it faces exactly the same problems as concerns scope and the *Speciesist Scientist* as the bias variation above. Because it is still restricted to moral worth it is unable to account for counter-examples based on other traits, and because it remains a mental-state theory it is forced to implausibly claim that Researcher 1 does more wrong than Researcher 2. The situation is only marginally better for the challenge of misestimation, as all the added condition does is

specify that any estimation of particular values, irrespective of accuracy, is morally wrong. But this still leaves the situation qua misestimates under the threshold unresolved.

On top of this, the opacity variation appears to have no way of explaining any wrongness pertaining to discrimination stemming from estimates over the threshold. Suppose the *Bigoted Billionaire* discriminated based on his taking into account values of the pertinent traits above the threshold. Since he takes these values into account for both the favoured and the disadvantaged group, it seems he must be held to be treating both groups disrespectfully. But we would like to say that in discriminating this way, he wrongs the discriminatees – those disadvantaged and discriminated against – in a way he does not wrong the favoured group. The opacity variation seems incapable of accounting for this difference.[34]

Let us turn therefore to the next variation, which introduces instead an element of appraisal respect.

---

[34] A proponent of disrespect might now object that we have not actually considered a genuinely negative account. The opacity-account is only partially negative, in that it introduces a cap on the qualities of certain traits that must not be taken into consideration, but it also contains the positive requirement of taking moral worth into consideration. It is a requirement, that is, to take all and only moral worth – and the constitutive traits in so far as, but only in so far as they are constitutive – into consideration. It is perhaps better described as a combined account. Now, while I have no desire to adjudicate among these competing versions of disrespect (positive, combined, negative), fortunately I do not believe that I need to. A straight-forward negative account runs into many of the problems we have seen above. I illustrate this in discussing the prejudice-animosity variation below.

The valuing variation: disrespect from failure to properly appreciate

In a recent *Ethics* article, Joshua Glasgow presents disrespect as potentially useful in the conceptualisation of racism in a way that runs parallel to the discussion I am here pursuing.[35] (Glasgow, 2009) Glasgow's motivation for looking at disrespect is that attempts to provide a comprehensive definition of racism flounder on what he calls 'the location problem', that is, the apparent difficulty in determining whether racism is cognitive, behavioural or attitudinal. Attempts at locating the root of racism decisively in any one location run into credible counterexamples based on one of the other two: "In short, we call beliefs 'racist' even when they neither issue in racist behavior nor issue from racist noncognitive attitudes; we call attitudes 'racist' even when they fail to effect racist behavior and are unaccompanied by racist beliefs; and we call some behavior 'racist' even when it takes place in the absence of racist beliefs or attitudes." (Glasgow, 2009, p.69) Glasgow's solution is to adopt the "Disrespect Analysis" of racism: "(DA) φ is racist if and only if φ is disrespectful toward members of racialized group R as Rs." (Glasgow, 2009, p.81)

---

[35] Glasgow's focus is therefore slightly different, and his stated aim is not to defend the disrespect-account but to "propose a unified account of racism".(Glasgow, 2009, p.64) Nonetheless, the central ideas and arguments that Glasgow presents in the pursuit of his unified account apply, I believe, to discrimination in the wider sense, and his unified account involves, indeed hinges on, a qualified defence of the disrespect-account.

At this point, we want to know what Glasgow understands by 'disrespect'. Indeed this is particularly urgent for Glasgow because, as he rightly notes, the most obvious challenge for his disrespect-account is that it must hold disrespect to be different from simply an attitude or return full circle to the location-problem. He initially suggests that we may be able to rely on shared intuitions about the common usage of 'disrespect': "Instead [of giving a full theory of disrespect], I hope that most readers, whatever their theoretical persuasions, have some sort of pretheoretical grip on respect and can agree on a series of commonly recognized cases of disrespectfulness: systematic suppression of Rs' political rights, workplace discrimination against Rs, the utterance of racial epithets, hating Rs as Rs, and so on." (Glasgow, 2009, p.86) And he confronts the attitudinal-challenge by arguing that disrespect can be used to describe a variety of phenomena in all three locations: "We say things like: "Your dismissive attitude is disrespectful," "Your claim that Kerry is a coward is disrespectful, particularly in light of his exemplary military service," and "Giving him 'the finger' was disrespectful." Indeed, we say such things even when the attitudes, beliefs, and behaviors seem to come apart from each other. We might tell a child who "gives someone the finger" just because that's what he saw his older brother do to stop doing that on the grounds that it is disrespectful, even though the child had no disrespect in his heart. Thus it appears that not all behavioural or cognitive disrespect derives from attitudinal disrespect." (Glasgow, 2009, p.83-84) Moving from racism to the wider context of discrimination, this would suggest that an agent

engages in *broadly disrespectful* morally wrongful discrimination against a patient *iff*:

> 1) she treats T-persons differently than ¬T-persons, and
>
> 2) the treatment is disadvantageous for T-persons,[36] and
>
> 3a) the attitude or belief that explains the differential treatment can be called disrespectful towards the discriminatee, or
>
> 3b) the actions of the discriminator constituting her treatment of the discriminatee(s) can be called disrespectful towards the discriminatee

There are several problems with this account, however. First, I am probably less comfortable than Glasgow is putting my trust in 'ordinary usage'. The concept of respect, it seems to me, is by virtue of its abstract and normative nature one that we might expect to be both confused and contested in public discourse.[37]

---

[36] Note again that, although Glasgow stresses that even potentially beneficial and seemingly benign paternalist forms of racism can be morally wrong, I still want to say that discrimination *against* requires that the discriminatee is somehow disadvantaged. In effect, Glasgow's claim amounts, as does the disrespect-account in general, to the claim that discrimination for and between can be as morally bad as discrimination against.

[37] As Robin Dillon aptly describes its complications, analyzing respect involves taking a stand on questions such as: "(a) What category of thing is it? Philosophers have variously identified it as a mode of behavior, a form of treatment, a kind of valuing, a type of attention, a motive, an attitude, a feeling, a tribute, a principle, a duty, an entitlement, a moral virtue, an epistemic virtue: are any of these categories more central than others? (b) What are the distinctive

Compare e.g. the pedestrian who 'respects' the law by stopping for a red-light at a crossing late at night with no traffic in sight, or 'respecting' the decision of a political figure to do something controversial in her private life, such as divorcing or converting her religion, to the surfer who 'respects' a dangerous wave by being extra careful when riding it or 'respecting' the achievement of a mountaineer who scales a difficult peak. As I have tried to show, it is not only that there may be cases where we can disagree about whether or not we ought to respect something, but that 'respect' can mean different things in different contexts. So the mere fact that the *term* 'respect' (or 'disrespect') could be or is used to describe and evaluate both beliefs, attitudes and actions is no guarantee that three identical *phenomena* of respect are at work. To ensure this, we would at the very least have to try to restrict the range of cases to those where respect has moral connotations.

---

elements of respect? (c) To what other attitudes, actions, valuings, duties, etc. is respect similar, and with what does it contrast? (d) What beliefs, attitudes, emotions, motives, and conduct does respect involve, and with what is it incompatible? (2) What are the appropriate objects of respect, i.e., the sorts of things that can be reasonably said to warrant respect? (3) What are the bases or grounds for respect, i.e., the features of or facts about objects in virtue of which it is reasonable and perhaps obligatory to respect them? (4) What ways of acting and forbearing to act express or constitute or are regulated by respect? (5) What moral requirements, if any, are there to respect certain types of objects, and what is the scope and theoretical status of such requirements? (6) Are there different levels or degrees of respect? Can an object come to deserve less or no respect? (7) Why is respect morally important? What, if anything, does it add to morality over and above the conduct, attitudes, and character traits required or encouraged by various moral principles or virtues?" (Dillon, 2009)

174

Glasgow, however, at first appears to be willing to bite this bullet: "…if 'disrespect' itself connotes a moral negative (as I believe it does), then even if all instances of disrespect have nothing else in common besides that moral valence, they will all be at least defeasibly morally condemnable, and in that case DA can enable the normative work we want done by an analysis of racism." (Glasgow, 2009, p.88) Once we turn to the cases involving moral valence however, disrespect may be too thin a concept in much common usage for it to be useful. My feeling is that in common parlance disrespect is frequently a non-substantive pejorative, in effect synonymous with 'wrong', 'bad', 'impermissible', 'disagreeable', 'undesirable', 'unpleasant', etc. with few or no implications pertaining to that particular term being applied rather than one of the others. In such cases, if not as emotivism would hold in all, 'disrespectful' really does just mean 'boo!' (cf. Ayer, 2002 [1936]) Taken together, these two concerns substantially undermine Glasgow's reliance on a common-sense concept of disrespect. It is the vagueness of the concept which allows its broad normative scope, but we will only have common agreement on such cases if we use disrespect in the broadest sense possible, i.e. the one where it is the equivalent of booing.[38]

---

[38] Note two additional implications here. First, the position Glasgow adopts is inconsistent with the project first outlined, in that it represents a retreat into the approach where we define racism by convention simply as those varying cases where race-based discrimination is for any reason morally condemnable. Now as noted initially this is an easier position to defend, but it is so because it gives up explanatory power. Secondly, the resort to intuitions that Glasgow here relies on

Ultimately, unless we have some understanding of what it means for something to be disrespectful, there is no way to know whether or not any overlap is accidental, or indeed whether there is any overlap between the two concepts of racism and disrespect. Glasgow perhaps concedes as much when latter in his article he moves closer to the concept of recognitive respect, as he admits to being "…partial to understanding the relevant kind of disrespect as something like a failure to adequately recognize autonomous, independent, sensitive, morally significant creatures."[39] (Glasgow, 2009, p.86) Further clarification, unfortunately, is brief and relegated to the footnotes, where he identifies his position with Darwall's recognition respect, but with two modifications.[40] First,

---

seems to me incapable of doing the conceptual work that he needs it to: Even if we accept that disrespect connotes negative moral valence, and that hence we should only call something disrespectful if it is morally condemnable, the issue at stake is whether these cases fit the broad range of common usages that Glasgow suggests. Unless we are willing to take *all* uses of disrespect in ordinary parlance as valid, we need a thicker description of disrespect to distinguish between the cases it does apply to and those it does not.

[39] Note that Glasgow demonstrates familiarity with the challenge I described above in the context of the *Speciesist Scientist*: "The "overattitudinalization" of disrespect seems to have created some unnecessary image problems for respect-based accounts. For example, Kasper Lippert-Rasmussen [...] rejects respect-based accounts of discrimination (rather than racism) because he thinks they entail that discrimination can be bad only when someone is "actually being disrespectful" and that the only way out of this problematic entailment is to have the discriminator falsely represent the discriminatee as having a lower moral status. As I will argue, we need not take DA down either of these paths." (Glasgow, 2009, p.83, note 44) As far as I can tell, Glasgow does not make good on this promise, although his suggested modifications (see below) could perhaps be taken to represent his proposed alternative path.

[40] Rather than present a concise definition Glasgow excuses himself by reasonably suggesting that the task of developing a full account of disrespect would exceed the scope of his article, although one is tempted to object that this

he favours the 'valuing recognition' proposed by Robin Dillon: "…at least sometimes, the appropriate recognition will involve valuing the object of respect, when the object is a person." (Glasgow, 2009, p.86, note 49; Dillon, 2009) It far from clear, however, how Glasgow takes valuing recognition to be a modification. Recall that (positive objective) *recognition respect* obtains *iff*:

> 1) the agent recognizes the relevant feature of the object of respect,
> 2) gives this feature the appropriate weight in her deliberations, and
> 3) acts in accordance with what possession of the feature requires

whereas *appraisal respect* obtains *iff*:

> 1) the agent recognizes the extraordinary quality of the relevant feature of the object of respect, and
> 2) adopts an attitude of positive appreciation for the feature

Now, according to Dillon: "…valuing is essential to some forms of respect that are not appraisal respect. In particular, valuing persons intrinsically is widely regarded as the heart of the respect that all

---

is so only if he truly needs to develop a new and alternative account rather than refer to or modify an existing account of respect.

persons are thought to be owed simply as persons. However, it is not sufficient simply to gloss recognition respect as recognizing the value of the object, for one can recognize the value of something and yet not value it, as an insurance appraiser does, or take the value of something, say, a person's child, into account in deliberating about how best to revenge oneself on that person. Respect for some categories of objects is not just a matter of taking the object's value into consideration but of valuing the object, and valuing it intrinsically." (Dillon, 2009; cf. also Dillon, 1995a, , particularly p.14-21) It is not entirely clear to me to what extent Dillon considers these views to be contrary to Darwall's or a clarificatory reiteration, but they seem to me to be largely in agreement with his understanding on central points.[41] Of the two forms of respect Darwall defines, it is appraisal that is concerned with valuing in the sense of having an attitude of appreciation. But Darwall's definition of recognition respect is clearly normative, in the sense that giving a morally relevant feature 'appropriate' weight in one's deliberations precludes e.g. estimating its usefulness in achieving some illicit purpose, on any plausible moral theory. This is either giving it weight, but inappropriate weight, or responding to (and evaluating) the wrong feature altogether – say, the child's emotional value to her parents rather than her moral worth as a person. So, if on the one hand valuing means *appropriately* recognizing and weighing the value of some feature (such as

---

[41] Dillon argues elsewhere for the inadequacy of a notion of respect focused on the moral worth shared by persons, but this, I take it, is a separate issue. (Dillon, 1995b)

personhood), this is already a central feature of our definition. If on the other hand it means adopting a particular attitude, then it introduces the substantive normative claim that one of the ways we are obliged to act when confronted with certain things of value, persons in particular, is to adopt attitudes of positive appreciation. Let us suppose, given that we are looking for a possible alteration of Darwall's position, that this is the interpretation that Glasgow has in mind.

Building on Dillon we might then say that an agent is (positive objective) *valuing respectful* of a person *iff*:

> 1) the agent recognizes the relevant feature of the object of respect,
> 2) gives this feature the appropriate weight in her deliberations,
> 3) adopts an attitude of positive appreciation for the object's possession of the feature, and
> 4) acts in accordance with what possession of the feature requires

Based on this definition we ought, it seems, to say that an agent engages in *valuing disrespectful* morally wrongful discrimination against a patient *iff*:

> 1) the agent treats T-persons differently than ¬T-persons,
> 2) the treatment is disadvantageous to T-persons, and

3) the differential treatment is caused by a) her failure to recognize and appropriately weigh the moral worth of T-persons, or b) her failure to adopt an attitude of positive appreciation for the moral worth of T-persons[42]

How well does this idea stand up to scrutiny? Unfortunately, not very well. First, it too fails to do better than the bias variation with respect to the three basic challenges. It gives us no better grounds for selecting an account of misestimation, it fails to include large groups of plausible examples of wrongful discrimination, and if

---

[42] Alternately, one might want to say that the demand of appraisal respect (3b) replaces, rather than supplements, the existing condition (3a). Glasgow, when discussing his second suggested alteration, seems to favour this approach and abandoning "the deliberative focus of Darwall's analysis": "Even putting institutional respect to the side, it is arguable that these words should be removed anyway, since our attitudes seem capable of manifesting valuing recognition respect or disrespect whether or not they figure in our deliberations. And bringing institutions back in, it seems fair to say that both institutions and agents can engage in valuing recognition respect. Even if they have no attitudes, institutions and their policies can (and should) reflect our equal moral status." (Glasgow, 2009, p.86-87, note 49) His solution is the shift to "acting in a way that reflects equal moral status". I find his arguments to that effect unpersuasive, since equal oddity attaches to institutions 'appreciating', and since the obscurity of what it means to 'reflect moral worth' robs the definition of explanatory power. It is also questionable whether the apparent problem which motivates Glasgow to make the alteration is well conceived. A better solution might be to admit that when we speak e.g. of institutions or laws being respectful we are speaking metaphorically, in that while it is impossible for them to literally deliberate or maintain the attitudes required to be respectful they can manifest virtues that are relevantly similar to what we mean and approve of when an agent is respectful. (cf. Lægaard, Forthcoming) In any case, substituting 3b) for 3a), rather than supplementing, makes no substantial difference to the problems I outline below.

anything it seems to strengthen the contrary intuitions of the *Speciesist Scientist* if we assume that the unbiased researcher has a positive appreciation appreciation for the moral worth of the lab-animals.

Furthermore, it introduces a series of problems of its own. The first problem is that a duty of appreciation is probably better identified with specific moral interpretations of the substantive normative claims that moral patients can raise than it is with the concept of (recognition) respect itself. This is the point observed by Darwall when he underscores that: "…what [recognition respect] requires as appropriate is not a matter of general agreement, for this is just the question of what our moral obligations or duties to other persons consist in." (Darwall, 1977, p.38) Thus, there is at least a disputable issue of whether the duty to adopt an attitude of positive appreciation is best understood as part of what follows from recognizing and weighing the moral worth of persons, i.e. a consequence of recognition respect, or as part of what being respectful means. Proponents might cede this point and maintain that a duty of appraisal respect is nonetheless a central duty that follows from moral worth. But to do so means that it must be defended independently of the claim that we have a duty to be respectful. We have granted for the sake of argument that there could be a duty to recognition respect, but this does not include an assumption that a duty to appraisal respect follows from the recognition of moral worth. Of course, this is not to suggest that defending this claim is an insurmountable challenge. Only that it does place an extra burden on the argument.

A second, and probably more serious problem, is that appraisal respect does not seem to mesh well with our intuitions about what is wrong when something is wrong with discrimination. Consider:

> *Sophie's Choice.* Sophie the saboteur is faced with a grim decision. Having planted a bomb on a deserted square to destroy military vehicles parked there, she suddenly sees two groups of innocent civilians enter the square from opposite sides. Suppose that although she has reason, of course, to save both groups, she has time to alarm and save only one of the groups, and that one is bigger than the other. Suppose further that, as a jaded agent of the resistance, Sophie still recognizes and acts in accordance with the moral worth of persons, but she suppresses her feelings, including feelings of appreciation, because these would often render her incapable of acting. Suppose finally that Sophie saves the bigger of the two groups.

It seems clear that in choosing to save one group, Sophie treats the two groups differently by distributing a disadvantage onto the other, and so discriminates against the group she does not save in the instrumental sense I defined initially. It also seems to me that this is not a case of wrongful discrimination. Sophie was about to

kill both groups, and made the morally right decision to save the one group she had most reason to save.[43] But, it also seems that if it is counterfactually true that Sophie would have been incapable of acting at all had she felt appreciation for the moral worth of the persons in the two groups, and would therefore have saved neither group, then the valuing disrespect-account is implausibly forced to claim that *Sophie's Choice* is in fact a case of wrongful discrimination!

In conclusion, the valuing variation seems as unsuccessful as the opacity variation. Neither of the two alterations of the original bias account avoid the challenges facing the original, and they each introduce new problems. Let us turn therefore to two accounts that make more radical breaks with the premises of recognition respect.

The expressive variation: disrespect as demeaning

An interesting alternative understanding of the disrespect-account is developed in Deborah Hellman's "When is discrimination wrong?" (Hellman, 2008). Focusing on what we may call an

---

[43] As Sophie is directly causally responsible for the deaths of those she does not warn (she has planted the bomb), I think it is at least arguable that she does not merely "let them die", a distinction that at least some proponents of disrespect-accounts would probably want to hold is important to how we should assess this scenario. Even if she merely allows the deaths of one group, she may still have moral grounds for discriminating. Some might then want to hold that these grounds are based on something other than the size of the groups, such as e.g. the fair outcome of a lottery (e.g. flipping a coin). I think the case can easily be modified to accommodate such concerns.

expressive account of disrespect, Hellman potentially avoids the basic challenges by shifting the location of the wrong of discrimination to the meaning of the discriminatory act. The variations that I have discussed above are mental-state principles which locate respect in the deliberations of the agent, and take respect to be the primary concept with disrespect defined primarily as the negation of the respect. Hellman's account locates disrespect in the intersubjective social space of expressions, and takes disrespect to be the primary concept with respect defined primarily as the negation of disrespect.

Hellman initially declares that she considers the equal moral worth of all persons to be a "bedrock moral principle", and that the she takes it to consist in that: "…all people are equally important from the moral point of view and so are equally worthy of concern and respect." (Hellman, 2008, p.6) This ties into wrongful discrimination on her view in that explaining the wrongness is essentially to answer the question: "… when does drawing distinctions among people fail to treat those affected as persons of equal moral worth?" (Hellman, 2008, p.7) But what does it mean to fail to treat the affected as persons of equal moral worth? Hellman focuses on the idea of 'demeaning': "What one *does* in drawing a distinction on the basis of some characteristic is not just separate people into two or more groups and allocate different treatment on the basis of that distinction. Sometimes one also demeans some of the people one classifies." (Hellman, 2008, p.24-25) And, Hellman concludes in a central section: "Demeaning actions are those that put the other down. To demean is to express

that the other is less worthy of concern and respect and to do so in a manner that has power."[44] (Hellman, 2008, p.57) This at least suggests *when* we "fail to treat persons as moral equals" in the relevant sense – namely when we act in a way which 'expresses' that they are not – but leaves the question of why, ultimately, demeaning is wrong unanswered. Hellman's answer is simply that: "Demeaning is wrong because the fact that people are of equal moral worth requires that we treat them as such. We must not treat each other as lesser beings even when doing so causes no harm." (Hellman, 2008, p.30)

---

[44] Unfortunately, although the concept is explored at some length, this is about as close as we get to a strict definition of demeaning, perhaps because Hellman understands her use of the term to be sufficiently close to common usage that no definition is necessary, as when she suggests that: "To demean is not merely to insult but also to put down, to diminish and denigrate. It is to treat another as a lesser." (Hellman, 2008, p.29) At other times the definitions she suggests appear circular, as when we are informed that "[to] demean is to treat another as not fully human or not of equal moral worth." (Hellman, 2008, p.35) A non-circular reading of this would imply that Hellman is stipulating rather than defining, so that 'demeaning' is simply the term she will apply to that, as yet undefined group of actions which "treat another as not fully human or not of equal moral worth". Further, Hellman distinguishes demeaning from merely expressing disrespect. "…demeaning requires an especially strong expression of disrespect – that of a lack of respect for another's equal moral worth." (Hellman, 2008, p.36) Disrespect, it seems, is a necessary but insufficient condition for demeaning. Since it is not exactly clear what she believes expressing disrespect to be – whether e.g. it is stating a belief or a dislike – this does not clarify matters much, even if she does mention as examples of expressing respect "taking off one's hat when entering a room, writing a thank you note to one's dinner host, looking someone in the eye when speaking", and of expressing disrespect "giving someone the finger, spitting on someone, looking over someone's shoulder when she is speaking to one". (Hellman, 2008, p.36)

Three points about this tentative definition are worth emphasizing. First, Hellman stresses that demeaning differs from stigmatizing in that it is not defined by its impact: "Rather than emphasize the effect (psychological or social) produced by classification, I claim that sometimes it is wrong to classify because of what one expresses – regardless of whether the person or people affected feel demeaned, stigmatized, or degraded." (Hellman, 2008, p.27)

On the other hand, Hellman holds that demeaning does depend at least partially on the status-relations of the involved parties, in that a discriminator must possess a certain amount of influence over the discriminatee for the discrimination to be demeaning: "To demean is to put down – to debase or degrade. To demean thus requires not only that one express disrespect for the equal humanity of the other but also that one be in a position such that this expression can subordinate the other."[45] (Hellman, 2008, p.35)

Thirdly, it is no coincidence that she describes demeaning as 'expressing' that the discriminatee is inferior. As becomes clear during her discussion of how to determine the wrongness of discrimination, Hellman holds that it is the nature of the action

---

[45] Hellman explains this distinction through an example that holds that spitting at a colleague or superior will typically not demean, while spitting at a homeless person very likely will, the difference being that only the latter is liable to "put down" the victim. (Hellman, 2008, p.35) Since this is not to be understood as an effect subject to the victim's understanding, such as e.g. the harm to self-esteem which might well result in the homeless scenario, but not, or at least not to the same extent, in the colleague-scenario, I confess that I am uncertain what it is meant to be.

which is at stake, and that the central task is therefore an interpretative one of finding out whether the action objectively interpreted is disrespectful by *expressing* that the discriminatee is morally inferior.[46] (Hellman, 2008, p.59-85)

Summing up we might say that that an agent wrongfully discriminates against T-persons in an *expressively disrespectful* way *iff*:

> 1) the agent treats T-persons differently than ¬T-persons,
> 2) the treatment is disadvantageous to T-persons, and
> 3) the differential treatment 'demeans' the discriminatee, i.e. given the context and roles of the discriminator and discriminatee the treatment expresses that T-persons do not have moral worth equal to ¬T-persons

The expressive variation may appear promising, because arguably it fares somewhat better with respect to one of the basic challenges. Thus, Hellman avoids running counter to intuitions in the *Speciesist Scientist* the way earlier variations have, because although she is still

---

[46] Hellman's discussion of what it means for an act to express the moral inferiority is itself less clear than one might wish for. Her notion of interpretation seems partly conventional and partly counterfactual, thus it relies both on the traditional meaning of certain actions, such as racial segregation, and on a kind of assessment of whether the agent acts 'as if' she discriminated on the basis of differential moral worth. Neither approach, it seems to me, can provide anything like the 'objective' interpretation Hellman desires. While I consider this vagueness a weakness in her argument, I shall set it aside.

required to hold that contrary to what some might feel researcher 2 does no more wrong than researcher 1, she is not required to hold the even more counterintuitive claim that researcher 1 does more wrong than researcher 2. Both researchers, if we assume that they demean the animals they discriminate against and experiment on, do equal wrong.

As for the limited scope of her account of wrongness, Hellman explicitly bites the bullet, which while not a solution to the problem at least suggests that it is a cost which Hellman considers acceptable. Whether we ought agree with her is questionable. Imagine the following variation on the *Bigoted Billionaire*:

> *Prejudiced PC.* The (unbigoted) billionaire purchases the company, but rather than firing workers with one or more particular traits, he uses a computer programme that considers a variety of data and picks workers randomly, although while assigning slightly greater chances of being picked to workers who are at the bottom of a ranking for certain performance-indicators. Suppose that this procedure is fair and non-discriminatory. However, unbeknownst to all parties involved, the software contains a programming error which makes it select only from a group defined by a particular trait, such as race or gender. All the randomly selected workers will be drawn from this sub-group.

Suppose that nobody notices this, as all parties believe the programme to work.

Is the discrimination here morally wrong? Hellman, it seems, is left with two unattractive options. She can either claim that since nobody acts in a way that expresses disrespect, the disadvantageous treatment does not in fact constitute wrongful discrimination. Or she can claim that although nobody understands the treatment to express disrespect, it does in fact demean the workers and the wrong does in fact consist in just this, rather than the more obvious alternative of consisting in the way that the treatment of the workers substantially fails to treat them as equals by assigning them unequal chances of carrying a cost.

It is even less obvious that the expressive account can avoid problems parallel to those that affect misestimation, although it is the representation of moral worth that is at stake. Note that the misrepresentation of moral worth in Hellman's account is explicitly comparative. Her concern is not the accurate evaluation of moral worth, but the accurate representation of comparative equality of moral worth. Hellman's argument for her concern with comparative expressive disrespect emerges in a discussion of Harry Frankfurt's critique of egalitarianism. (Frankfurt, 1997) As Hellman notes, Frankfurt, although strictly speaking not directly concerned with discrimination, does provide a classical disrespect-argument when he claims that: "Treating a person with respect means, in the sense that is pertinent here, dealing with him exclusively on the basis of those aspects of his particular character or circumstances that are actually relevant to the issue at hand.

Treating people with respect precludes assigning them special advantages or disadvantages except on the basis of considerations that differentiate relevantly among them. Thus, it entails impartiality and the avoidance of arbitrariness." (Frankfurt, 1997, p.8-9) To this traditional (negative recognition) respect-account Hellman objects that: "…this approach – saying that each person is entitled to the respect that being a person entails – is itself empty. How would one ascertain what treating someone with the respect appropriate to personhood requires? Rather, the fact that we all share a common humanity requires that we be treated *as worthy as* others. We give flesh to the injunction to treat others with the respect that our common humanity demands by saying that no one may be treated as a *second-class* person. In other words, there is something inherently comparative here." (Hellman, 2008, p.48) This seems problematic on several levels.

First, Frankfurt gives a plausible and classical answer to the question she poses, wherefore his account cannot be called "empty", even if Hellman may hold that it is mistaken: "The lack of respect consists in the circumstance that some important fact about the person is not properly attended to or is not taken appropriately into account. In other words, the person is dealt with as though he is not what he actually is." (Frankfurt, 1997, p.12)

Secondly, it is not clear that Hellman's criticism properly targets Frankfurt's argument. He is essentially concerned with criticising principles that attach intrinsic normative importance to equality of outcomes, which presumably Hellman would agree to consider morally irrelevant, and his interpretation of what equality of

treatment means differs from Hellman's, but not in the sense it appears that she believes. Thus, her concern with the difficulty of ascertaining "what treating someone with the respect appropriate to personhood requires" seems as pertinent to her own account as it does to Frankfurt's, and her answer – that it requires not acting in a way which expresses disrespect (the belief that the person affected is morally inferior) – warrants a normative foundation as much as Frankfurt's.[47]

Thirdly, the claim that the concept is inherently comparative does not seem plausible for reasons similar to those outlined in the basic challenge. We would not consider responding to everybody in a way which failed to respect their common humanity to be morally acceptable, even if it meant treating each of them "as worthy as others". Rather, as Frankfurt points out, it seems that: "What most fundamentally dictates that all human beings must be accorded the same entitlements is the presumed moral importance of responding impartially to their common humanity" (Frankfurt, 1997, p.11) Impartiality however, which could well fit with Hellman's overall approach, is certainly not a comparative concept.[48]

---

[47] I elaborate on this immediately below.

[48] Note that I concur with Hellman that the principle of formal equality will not help us evaluate cases of potential discrimination. For that we need a substantive normative theory to indicate what the relevant features for which we should equalize are. Her mistake lies in believing that shifting to a procedural form of equality avoids this challenge. This kind of procedural norm must itself be justified by a normative principle. Oddly, in chapter 5 where many of the same

In addition to the continued difficulty with misestimation outlined above, there are two further difficulties introduced by the shift to expressive disrespect worth exploring. The first concerns what motivates the shift in focus. That is, why not simply say, as Hellman occasionally does, that "the fact that people are of equal moral worth requires that we treat them as such", i.e. that we recognize and appropriately take into account their moral worth? This is important, because although she seems to assume so, it is not clear that Hellman can ground one on the other. The steps in Hellman's argument appear to run something like the following:

> 1) All persons have equal moral worth;
> 2) having equal moral worth imposes a moral duty on agents to treat those that have it as being equally morally worthy;
> 3) actions can express meaning, irrespective of the agent's intention or the effects of the action;
> 4) discriminating can demean, i.e. treat the discriminatee in a way which expresses that she is not of equal moral worth.

---

issues emerge, Hellman fails to follow up on the discussion here, or apparently to see their relatedness. Her statement, that she considers that "treating people as moral equals requires that one have some reason (and not a patently bad reason) to draw distinctions among them and thus that arbitrary differentiation is morally wrong", fits snugly with Frankfurt's declared stance, yet she does not consider his arguments to this effect. (Hellman, 2008, p.90)

As far as I can tell, Hellman takes this to demonstrate her conclusion, but of course the above premises do not entail:

> 5) when discrimination demeans, it is morally wrong (QED)

Crucially, this is because we have not established that expressing that a person is not of equal moral worth is a way of *treating* her as not being of equal moral worth.[49] Consider the following scenario:

> *Dutiful Demeaner.* An agent must impose a cost on either A or B. Imposing the cost on A will make little difference to A's situation, but will demean B, whereas imposing the cost on B will greatly harm her but not demean A. Suppose also that there are no relevant duties apart from the general duty not to cause undue harm, so that it is not the case e.g. that A deserves to pay the cost more than B. When deliberating, the agent considers the two to be morally equally worthy, so that the much greater harm to B clearly establishes a *prima facie* case for imposing the cost on A. Assume finally that, consistent with Hellman's definition, demeaning does not depend on its effects, and that in this case

---

[49] That is, 2) and 4) do not entail the implicit premise that "it is morally wrong to demean a person".

demeaning B will not in fact harm her in any
way.

In this case, treating B as morally equally worthy seems to require demeaning her, although doing so will arguably violate an independent duty to not demean her. But if this is true, then demeaning and treating as morally equally worthy are partially coextensive, and Hellman requires an independent argument for the wrongness of expressive disrespect.

Secondly, in addition to the fact that respect for equal moral worth cannot ground Hellman's account of expressive disrespect there is good reason to be sceptical of whether Hellman's account of expressive disrespect can be independently grounded. What reason is there to consider the expression of a mistaken belief in unequal moral worth to be wrong, in and of itself? Hellman has excluded most of the ways of doing so, because, although context and status is obviously important for the impact that a message can have, Hellman has explicitly ruled out any reliance on the effects of the action such as stigma or other harm.[50] Nor does it seem plausible to suggest that demeaning somebody could actually change the moral status of a person. Expressing that someone as morally

---

[50] Indeed, Hellman often writes as if what matters is the potential harm of expressive acts, such as when in arguing that demeaning is sufficient to establish moral wrong, she holds that "avoiding demeaning treatment may [...] be a *central* interest of people." This, it appears that she accepts, is so because: "...self-respect, surely a central interest, is inextricably tied to avoiding demeaning treatment." (Hellman, 2008, p.48-49) Interests, and the interest of self-respect, are surely related to a harm-account, however.

194

inferior does not *make* them morally inferior. Hellman wants to place the locus of the wrongness in the expression, not its consequences, but consider how strange this idea is – why should we hold misrepresentation of human dignity to constitute a moral wrong, independently of its consequences? After all, we do not take most cases of misrepresentations to constitute such wrongs; "[i]t is not generally", as Nir Eyal puts it, "the case that the truth of p makes the expression of $\sim p$ a serious moral offence." (Eyal, 2003, p.24) Thus, we ordinarily assume that it is not in general wrong to express untrue beliefs. In fact, many people are mistaken about a great many things, but we do not normally find it morally problematic in the way discrimination is that they express their mistaken beliefs. On the contrary, we would normally hold that we need extraordinarily strong countervailing reasons to consider mere expressions of opinions or beliefs to be morally wrong, and that those reasons can normally only take the shape of harms caused by the expression, an explanation that Hellman has explicitly ruled out.[51]

In conclusion, although perhaps partly due to underspecification, Hellman's expressive disrespect-account does not appear to be a plausible explanation of what it is that is wrong when something is

---

[51] An exception might be that it is wrong to lie irrespective of the harm caused, but that seems a more complicated and narrowly circumscribed phenomenon involving both that the utterer does not believe the expressed belief and the intention of deceiving.

wrong with discrimination. Let us turn therefore to the last, and probably the strongest of the four variations.

The culpability variation: disrespect from unwarranted animus or prejudice

Richard Arneson, in a 2006 article, approaches the topic of discrimination from an angle similar to that with which we have been so far concerned, but offers an interesting variation on the explanation of the basis of the wrongness. Thus: "Discrimination that is intrinsically morally wrong occurs when an agent treats a person identified as being of a certain type differently than she otherwise would have done because of unwarranted animus or prejudice against persons of that type." (Arneson, 2006, p.779) Animus is defined as "...hostility or, more broadly, a negative attitude, an aversion", and prejudice as "beliefs about the person's characteristics that are either inferred from one's beliefs about persons of that type or directly caused by one's reaction to the type, these beliefs being formed in some culpably defective way." (Arneson, 2006, p.787, 788)

Arneson's account differs from the bias-variation in several respects then. Firstly, animus seems to me essentially a form of (objective negative) appraisal (dis)respect. Secondly, prejudice seems a form of (objective negative) recognition (dis)respect, but one which accepts a much broader range of mistaken beliefs than

simply beliefs about moral worth.[52] This has the benefit of alleviating the basic challenge of too narrow a scope – the culpability variation will be capable of encompassing cases such as the one I described where the belief pertains to laziness and stupidity rather than moral worth.

Thirdly, and most interestingly, the culpability variation introduces a constraint, which narrows the range of mistaken beliefs along a different dimension. This distinction emerges as the difference between what Arneson terms merely 'defective discrimination' and morally wrongful discrimination. Defective discrimination is, on Arneson's account, itself insufficient to establish moral wrongness: "One person may fail to respond to another in the right way given the circumstances, or respond by treating the other in ways that fail to adequately respond to the reasons that dictate how the other ought to treated, without the failure amounting to wrongful discrimination. The extra bit that when added to generic defective discrimination constitutes wrongful discrimination is the fact that one is led to defective conduct toward the other by *unjustified* hostile attitudes toward people perceived to be of a certain kind or

---

[52] The culpability variation is somewhat similar, in these two respects, to a combination of the opacity and valuing variations. Indeed, Arneson is perhaps motivated to widening the scope in this manner by concerns similar to those described in the opacity variation. He has argued elsewhere that there is reason to be sceptical of the claim that there exists a morally relevant feature, such as moral worth, which is shared equally by all human beings. (Arneson, 1999)

faulty beliefs about the characteristics of people of that type."[53] (Arneson, 2006, p.779, my emphasis).

This suggests the following definition using the terms that I have applied so far: an agent engages in morally wrongful *animosity-prejudiced disrespect* discrimination against T-persons *iff*:

> 1) the agent treats T-persons differently than ¬T-persons,
>
> 2) the treatment is disadvantageous to T-persons, and
>
> 3) the differential treatment is caused by a) her culpable adoption of an unjustified attitude of disapproval for T-persons, or b) her culpable misestimation of the trait(s) possessed by T-persons.[54]

---

[53] Note also the condition that "these beliefs [be] formed in some culpably defective way" in the prior quotation. The use of 'respond' here is ambiguous, but, as I read Arneson's distinction, it is applied to both the level of practical reasoning and the level of action. Thus, I interpret the first 'respond' to concern recognizing and deliberating appropriately 'given the circumstances', whereas the second is clearly 'responding' in the shape of doing something, although in this case not the thing(s) that proper recognition and deliberation would show to be the right thing(s) to do. If this interpretation is correct defective discrimination meshes reasonably well with the definition of recognition respect that we have been using so far.

[54] We might alternately interpret being unjustifiably hostile towards or holding false beliefs about a person as itself being disrespectful, i.e. being a way of failing to properly take into account and be motivated by the salient moral features. This is closer to the initial formulation, but is, I believe, a less obvious reading of Arneson's full argument because he insists that the animus/prejudice-conditions is 'added' to already deficient discrimination.

In addition to its ability to include a broader range of cases, the most obvious benefit of the culpability variation is that it dodges the challenge of the *Speciesist Scientist*. Recall that the basic challenge points out that in plausible cases misestimation of moral worth seems intuitively – granting that it has any impact whatsoever – to alleviate rather than aggravate wrongness and that for this reason it is peculiar to suggest that it is the factor that makes a case of discrimination wrong. Arneson's argument concerning defective discrimination suggests that this may not be true if the agent is culpably responsible for the misestimation. Does this mean that we are finally dealing with a successful disrespect-account? Probably not, because the culpability variation still faces serious problems.

First, consider where the shift to culpable misestimation leaves the variation with respect to the *Speciesist Scientist*. It may go some way towards meeting the challenge, but although our intuitions may well be less certain I am not sure that they change sufficiently to establish the case for the variation. Consider this version of *the speciesist scientist*:

> A pharmaceutical company discriminates against animals, inflicting horrible pain on hundreds or even thousands of them, to provide a very small benefit to a small group of humans, e.g. the ability to buy and wear a new perfume without suffering a small risk of having a mild allergic reaction.
>
> *Researcher 1* misestimates the moral worth of animals, and therefore falsely believes that her

actions are morally permissible. She cannot be blamed for this misestimation.

Researcher 2 correctly estimates the moral worth of animals, and therefore correctly believes that her actions are morally impermissible. She pursues them regardless.

Researcher 3 misestimates the moral worth of animals, and falsely believes that her actions are morally permissible, because her perception is distorted by a) her unwarranted hostility towards the animals she conducts the experiments on, or b) her unjustified and false belief in some other state of affairs pertinent to the situation (such as the God-given role of animals in serving man). She is therefore culpable for her misestimation.

Scenarios one and two are simply restatements of the basic challenge. Scenario three is the variation. However, while it certainly seems plausible that scenario three is in some respects worse than scenario one – whether or not these are the relevant respects is a separate issue – it is not clear to me that scenario three is also worse than the real contender, scenario two. The intuitions of most people may be quite uncertain here, but I doubt that many will feel more secure than I in pronouncing scenario two the worst of the three, which is what is required to refute the challenge and establish the presence of culpable misestimation as an uncontroversial wrong-making feature.

If there is an argument to be had here, I take it to be that by adopting such attitudes or holding such beliefs, the agent has put herself in a position to respond inappropriately. For both animus and prejudice it is thus central that the agent ought not to hold the attitude or belief: "If one discriminates against dishonest persons on the basis of a warranted negative attitude toward those people or on the basis of accurate beliefs about the associated traits of those persons relevant to decisions as to how to deal with them, no wrongful discrimination occurs." (Arneson, 2006, p.796) One concern in this respect might be to what extent we can hold agent's responsible for their attitudes and beliefs; it is difficult after all to decide to believe or disbelieve some proposition, but Arneson holds that it is at least possible for an agent to carry responsibility, if the beliefs have been formed by a 'culpably defective belief formation process'. Such a process need not itself rest on prejudice or animus. It can in theory be the result merely of neglecting to exercise critical reasoning, if e.g.: "I am simply lazy in forming beliefs. I harbour no animus…but I discriminate…on the basis of negative beliefs…that I absorb from the prevailing culture. I do not subject these beliefs to the critical scrutiny that is epistemically warranted due to the general unreliability of popular beliefs." (Arneson, 2006, p.789)

The logic of this argument seems to require a modification of Arneson's definition on two accounts, however. If the wrongness adheres to the culpability of the agent for the disturbing factors which renders her liable to defectively discriminate, i.e. treat persons differently in a way that fails to respond appropriately to

their morally relevant features, then unwarranted *appreciation* and unjustified *true* beliefs both seem to qualify as grounds for wrongful discrimination on a par with animus and prejudice.[55] Consider, one more time, the case of *the bigoted billionaire*. To simplify things, let us suppose that the billionaire only discriminates on the basis of one trait, e.g. gender. And let us suppose that firing the group of workers, i.e. women, is a case of defective discrimination, so that it meets Arneson's minimum criteria. Now consider four variations:

> 1) The billionaire is led to act in this way largely by his unjustified, false belief that women are less desirable workers than men – less honest, less capable, less disciplined, etc.
> 2) The billionaire is led to act in this way largely by his unwarranted dislike of women
> 3) The billionaire is led to act in this way largely by his unjustified, true belief that women are less desirable workers than men – less honest, less capable, less disciplined, etc.
> 4) The billionaire is led to act in this way largely by his unwarranted appreciation of men

---

[55] Meanwhile, warranted attitudes and justified beliefs take us too close to the Alexander-account, and thus reintroduce the basic challenge in its full strength. That is, if the agent discriminates on the basis of a failed response to the moral features of the situation, but in a way that does not represent a culpable failing on her behalf, e.g. because she is justified in the beliefs that lead her to misestimating, then we are essentially back in scenario 1 above.

The first two variations are both and equally bad, according to Arneson. But in the third it is mere coincidence that makes the billionaire correct in his belief. We can suppose that he has no justification for his view, and has formed it through a culpably defective belief formation process. Arneson must, it seems, be committed to labelling this a case of wrongful discrimination, if we are to take seriously his notion that it is the intentions of the agent that matter. Similarly, but perhaps less surprisingly given the revision of Alexander's account, it seems very hard to explain why 4) (discriminating against women because of an unwarranted positive attitude towards men) should be different than 2) (discriminating against women because of an unwarranted negative attitude towards them).[56] These modifications need not be considered damaging in and of themselves to Arneson's version of the disrespect-account, although they will broaden the scope of the definition in a way that risks incorporating counter-intuitive examples.

An accompanying and perhaps even more problematic issue is whether in shifting the locus of the wrong-making to the culpability of possessing certain attitudes or beliefs, Arneson has not moved too far towards an extrinsic account of wrongness. Analogously, it is plausible both that causing a traffic accident that injures innocent people is wrong, and that drinking heavily before driving, thereby putting oneself in the position where one will

---

[56] Naturally, we need to keep the concern for the potential difference in harm caused by the two different types of attitude out of the disrespect-account.

cause the accident, is wrong. But is it also plausible that, on top of these two wrongs, drinking heavily somehow makes causing the accident even more wrong than it otherwise would be? Sifting through our intuitions here might be difficult, to say the least, because we need to somehow separate intuitions that pertain to the wrongness of putting oneself in a position where one should not be, e.g. because being there one poses a risk to others, and any intuitions that pertain to how being there transforms the wrongness of certain actions one takes because of being in that position.

Given the scope problem, the unresolved situation qua the basic challenge and the questionable intrinsicality of the wrong involved, I think it is fair to say that we may be better off looking for an alternative to this variation of the disrespect-account too.

Conclusion

In the course of this article I have sketched a raw picture of the normative concept of respect, introduced the disrespect-account of discrimination, reviewed five versions of it, and found all of them to suffer from a plethora of problems. The bias variation runs headfirst into the three basic challenges, which in combination seem to me to form an overwhelmingly strong case against it. The opacity and valuing variations do not manage to avoid the worst of these problems, and in addition raise problems with explaining the wrongness in cases of consistently taking differences above the threshold into account and explaining why an appreciative attitude is a feature rather than a consequence of

respect. Hellman's expressive variation is sufficiently different from the bias variation that it avoid some of the basic challenges, but has great difficulty specifying the grounds of its wrongmaking factor in an intuitively plausible way. To my mind Arneson's prejudice-animosity variation remains the strongest of the five. But the account is at least partially susceptible to the basic challenges, and it is not clear that it can locate wrongness in the background of the relevant beliefs without shifting to an extrinsic account of the wrongness of discrimination.

My conclusion is deliberately tentative. It is possible that one or more of the disrespect-accounts I have explored can be revised, expanded or clarified in ways which will improve its credibility. None of the problems I have outlined here represent, to my mind, a knock-down argument against the disrespect-account as a whole. But they do represent challenges that defenders of the account will want to take seriously. In the proud tradition of scholars everywhere, I conclude that at the very least much more philosophical work on the topic needs to be done.

Andrew Williams. I owe particular thanks to Nir Eyal and Jesper Ryberg for very helpful and detailed comments.

The Art of the Unseen – Three Challenges for Racial Profiling[§]

> "L'art de la police est de ne pas voir ce qu'il est inutile qu'elle voie." (Napoleon Bonaparte, Letter to Minister of Police Joseph Fouché, May 24, 1800)

Introduction: Controversy and Intuitions

"Profiling" can be provisionally defined as "the application of statistical evidence concerning differences in propensity for crime in a police practice when deciding whom to target." Racial profiling is thus the use of statistical evidence concerning racial differences in crime-rates in the attempt to increase the likelihood of apprehending offenders.[1] It is both in the public debate and the professional literature an enormously controversial topic, with strong opinions both for and against but not as much clarity as one might prefer about what *exactly* the pros and cons of racial

---

[1] How best to understand race is a both complicated and controversial issue. (cf. Corlett, 2011) In the following discussions of race, I do not mean to imply that there is a genetic basis for distinguishing between biologically distinct races, only that race is clearly part of the social landscape as a trait which is present in the perceptions of people, whether by identification of or self-identification with a race. (Loury, 2005) Furthermore, and perhaps more problematically, I shall use "race" as essentially synonymous with "ethnicity." This tracks the way the focus in the United States context tends to be "racial profiling," whereas in the European context it is "ethnic profiling," although the two are relevantly identical phenomena. Although this conflation is somewhat less conceptually refined than I should prefer, I do not believe that any accompanying problems affect my central arguments.

profiling are. In the following, I aim to explore these in some detail and sketch what I take to be three serious challenges for the proponents of racial profiling.

Let me begin with some fundamental observations. Much of the controversy regarding racial profiling stems, I believe, from the presence of two conflicting intuitions about the imposition and distribution of costs on those subjected to police practices. Consider the following scenario, a slight variation on a set of actual cases described by David Harris (Harris, 2002, p.210ff):

> *Chicago Customs.* You are returning from a visit abroad and re-entering your country at the airport. While going through customs you have your bags hand searched by a customs agent, carefully looking through every single item in your suitcase, backpack and pockets. You are questioned extensively about your travelling origin, purpose for travelling, destination, occupation, home address, etc. At the end you have to repack your luggage yourself. You are then directed toward a closed room off to the side. You ask for an explanation, but receive none. When you complain, you are yelled at by one agent, and the other remarks: "Oh no, not another one…" Once inside the closed, windowless room, you are placed hands against the wall, and searched thoroughly by the agents, who run their hands tightly over your entire

body including your genitals. When nothing is found you are strip searched, and when that too reveals nothing you are body cavity searched. Finally, having turned up nothing, the agents decide to apply a "monitored bowel movement," ordering you to imbibe a powerful laxative and wait for you to empty yourself into a conveniently placed non-flushable toilet. Only after several hours, when you have disclosed the contents of your digestive system to their content are you finally released. At no point do you receive an explanation for the treatment, and your explicit requests for legal aid and contact to your family, who expected you home a while ago, are consistently denied.

Most people would, I believe, be indignant and angry at being treated in the way described in *Chicago Customs*, and critical of a police practice which regularly and systematically subjected persons to such treatment. And rightly so. For although there are aggravating circumstances that could conceivably be remedied while maintaining the practice, such as the callousness of the agents handling your case, it seems clear that the bulk of your complaint concerns the discomfort, inconvenience and humiliation that you suffer as a consequence of the intensive scrutiny. This police practice imposes a heavy cost on its targets.

Now consider a different scenario, loosely based on data regarding stop-searches in London (London Metropolitan Police Authority, 2010):

> *London Metropolitan.* You live in an ethnically diverse neighbourhood with both Black, Asian and White occupants. Following a violent crime the police intensify their use of stop-searches, and one day on the street you are confronted by two officers who order you to submit to a search. They carefully search you by running their hands tightly against your entire body while you lean on a wall, legs spread and palms firmly on the bricks. A week later you are stopped again, and this time the officers escort you into an open gateway where, only partially shielded from public view, you are required to strip down to your underwear as they search your clothing. At the end, one officer pulls your underwear away from your body, while briefly shining a flashlight into it and checking your genitals. Three days later you face your third search, although this time only of the "pat-down"-variety you first experienced. Discussing the frequency with which you have been searched with friends, you find deep disparities along ethnic lines, and when you look up the statistics it turns out that members of your ethnicity have

at least a four times greater frequency of being searched than members of the dominant ethnicity.

Most of us would also, I think, intuitively find the police practice in *London Metropolitan* problematic. Surely, it cannot be right that the police discriminate against persons of a certain ethnicity to the extent that such persons are searched four (or more) times as frequently as others? Yet, and here is the rub, the intuitions I have attempted to entice in these two cases can, and often do, pull in opposite directions.

The imposition of such severe costs, in terms of e.g., inconvenience, discomfort and humiliation, as those described in *Chicago Customs* requires that the police have good reasons for imposing these costs. And such reasons it seems must consist in either a sufficiently great benefit in positively identifying a criminal – suppose that the police are searching for a smuggler carrying the final component necessary for a terrorist organization to arm and use a nuclear device – a sufficiently great likelihood that the persons subjected to the practice will be positively identified – suppose that the police know with complete certainty that someone matching a specific description will arrive with your flight and be carrying drugs, and that only you and one other person on your flight match the description, giving each of you a 0.5 probability of being the courier – or, ideally, a combination of the

two.[2] In *Chicago Customs* it seems to me clear that selecting targets completely at random, each of whom would have an equally and (presumably) very small probability of being positively identified as a smuggler of any kind, would be clearly wrong since the burden imposed on them is so great. In fact, the *only* way that the police can justifiably impose burdens of such calibre, given that they cannot alter the benefits of detecting a smuggler, is to carefully select for scrutiny only those persons who have a much greater than average probability of being smugglers.[3] Profiling is *necessary* to make such cost-severe police practices justifiable. Historically, what brought attention to the cases described by Harris was that they targeted black women, clearly engaging in both racial and gender profiling. Added to the wrong many felt was committed by discriminating in this way was the fact that hit rates for black women were lower than for comparative demographics, so that targeting this group seems to have been irrational in the narrow sense of *decreasing* the likelihood of successfully identifying drug smugglers. But the trigger for the outrage that followed is explained in terms of the fact that racial profiling was applied, whereas as I have stated above my initial feeling is that such practices would be unjustified no matter who they happened to,

---

[2] I shall return to and expand on why I consider this to be the case in the discussion of costs, benefits and justifiability below.

[3] This is perhaps putting it too strongly: "only" here is meant only to refer to potential justifications within the framework considered in this article, that is, consequentialist justifications. As per the note above, I expand on this in the pertinent sections below.

unless careful profiling meant that the practice had a high likelihood of apprehending criminals.

The pull of *London Metropolitan* is in the opposite direction. Although presumably any one of the searches described would have been uncomfortable the cost imposed is much less and therefore easier to justify than in *Chicago Customs*; it is their focused and cumulative effect that is deeply distressing. In the actual case of Metropolitan London, the monthly rates of stop-searches for the period of May 2009-May 2010 varied between 3.09 per thousand and 4.94 per thousand for whites, and between 12.89 per thousand and 18.38 per thousand for blacks. (London Metropolitan Police Authority, 2010, p.9) Asians, fluctuating between 5.42 and 8.34 per thousand, are much closer to the rates for whites, although still far more targeted.[4] What troubles us here, I think, is that the profiling targets particular ethnicities and we

---

[4] Figures from all of England and Wales suggest even stronger discrepancies between the various ethnic groups. Thus in 2008/09, which is the latest data available, the rate of stop-searches for whites was 18.6 per 1000, while that of blacks was a staggering 134.3 per 1000, a ratio of ca. 1:7. In both cases, these levels represent a steep increase in the rate of searches since 2004/05, but for whites the increase is 25.7% (up from 14.8 per 1000) while for blacks it is 55.4% (up from 86.4 per 1000). (Statistics on Race and the Criminal Justice System 2008/09, 2010, p.26; cf. also Phillips and Bowling, 2002) The 1984 PACE-regulation allows such stop-searches when there is "reasonable suspicion" that the person subjected to them carries contraband, e.g., stolen goods or illegal drugs. Note that the United Kingdom is, to my knowledge, the only European country that systematically collects data on the ethnicity of those persons who interact with the criminal justice system, and therefore the most readily accessible source for figures such as the one I cited. But although verifying this is thus difficult, I see no reason to believe that the U.K. is unique among its European neighbours in its profiling practices.

tend to assume that for police to discriminate along racial or ethnic lines is either intrinsically or instrumentally bad.[5] It might be intrinsically bad if e.g., police are obligated by considerations of justice to set aside all differences of gender, race and religion as irrelevant, that is, if justice requires the execution of the law to be "colour-blind." It might be instrumentally bad if, e.g., targeting vulnerable minorities will offend and stigmatize them, the imposition of costs upon them will further burden the already deprived or the cumulative effect of the burdens will cross a threshold beyond which the harm caused grows severe enough to matter morally in a different way. In any of these cases, the thrust of the intuition runs counter to the notion that profiling can be required for a police practice to be justified, in that it points out the ways that profiling can delegitimize an otherwise justified police-practice. Caught between the requirements of minimizing the imposition of costs on the innocent and of avoiding the

---

[5] Strictly speaking, we cannot infer that such discrimination is going on from the mere correlation between race and likelihood of being stopped. It is possible that both race and criminality correlate with some other trait, such as e.g., socio-economic deprivation, which is the trait actually profiled for. This would make it a case of indirect rather than direct discrimination. Further statistics, however, seem to indicate that police employ more profiling, and more prejudiced profiling, the more at liberty they are to do so: the ratio of black-to-whites in stop-searches under section 60 legislation, which has less strenuous requirements than "reasonable suspicion", increase to 16.9 to 1, while success-rates, i.e., stop-searches resulting in arrests, plummet from 12% to 4%. (Open Society Justice Initiative, 2009, p.34) In the following I shall not consider such complications, but focus the discussion on those cases where it is actually race or ethnicity that is being profiled.

unequal distribution of burdens onto racial minorities, where does this leave racial profiling?

The answer, I will suggest, depends on the balance of costs and benefits in a way not essentially different from how we ought to assess police practices more generally. However, contrary to what is frequently assumed, such consequentialist considerations do not support racial profiling, as a strong prima facie case can be made against it based on the relative certainty of the associated costs and the relative uncertainty of the associated benefits. Given this prima facie case against it the onus shifts onto proponents of racial profiling, but the same uncertainties which support the prima facie case make it unlikely that proponents will be capable of producing a solid counter to meet the prima facie challenge.

In the following, I first set out at some length the background for the discussion, including more stringent definitions of the central terms and a minimal account of the relevant costs and benefits and several versions of a principle of justification. Much of the discussion in this part of the article, the account of costs excluded, is of a general character, applying to police practices and profiling broadly. In the next section, I then suggest three challenges that racial profiling in particular encounters concerning its valuation, application and foundation. In conclusion, I argue that given these challenges and the difficulty of meeting them, racial profiling must be considered unlikely to be justifiable in realistic scenarios.

Policing and Profiling

Having briefly explored the background intuitions let me set out in more detail the central concepts of the discussion, specifically what it means to profile and how it relates to police-work more broadly.

By a "police practice," I understand roughly a limited set of actions performed routinely by police officers for a specific and pre-defined purpose. Patrolling a certain neighbourhood and stopping certain people to question and potentially search them to find drugs would be an obvious example. It consists of a set number of reasonably well-defined and routine actions which can be performed easily in much the same way by any police-officer ordered to perform the practice. But going to the house of a witness to question her, pulling over and testing a driver for inebriation or performing an arrest could also be understood to be practices. While leaving this definition deliberately broad, however, I want to emphasize an important delimitation that it is worth bearing in mind. Unlike the focus of the present article much if not most of the actual work done by police officers is not concerned with the apprehension of criminals. "Police activities," although different from one country to another, typically involve the ubiquitous paperwork that accompanies modern institutions, including such service-functions as the issuing of driver's licences, criminal records and passports, as well as e.g. traffic regulation, maintaining public order, community work, increasing perceived safety through visibility, accident control, and the many miscellaneous forms of "helping out" from returning lost children to their parents to making rowdy teenagers turn down the volume

of the stereo at house-parties. I set all these aside for present purposes, so that in the following when I refer to police practices, I shall in fact mean only those practices whose (immediate) purpose is to apprehend criminals. These include everything from the narrowly focused investigation of a serious crime to very broadly focused street-patrols intended among other purposes to allow officers to keep an eye out for any crime they happen upon. Of particular interest in the present context is the middle-ground between these two, occupied by the kinds of relatively focused screening exemplified by random alcohol-tests administered to drivers on specific days, in my native Denmark typically Fridays in December when there is traditionally a high frequency of drunks returning from company Christmas-lunches. It is in such relatively focused practices that profiling is most obvious and easiest to apply.

By "profiling" we normally understand the use by police of statistical evidence as a tool for identifying and apprehending criminals. The fact that it is merely one tool among others is important, because as Mathias Risse and Richard Zeckhauser have pointed out there is a tendency to mistakenly equate racial profiling with practices where race is the only or determinate trait governing which persons are subjected to the practice. (Risse and Zeckhauser, 2004, p.135-137) Thus with a slight broadening and rephrasing of a definition suggested by Risse & Zeckhauser we might initially say that profiling consists of "any police-initiated action that relies on statistical evidence and *not merely* on the

behaviour of an individual."[6] (cf. Risse and Zeckhauser, 2004, p.136)

I believe, however, that some further modifications are in order. First, note that the notion of statistical evidence at stake here is not rigorous. Officers need not know the exact ratio between male and female crime-rates to profile on the basis of gender; if they give special attention to male potential suspects based on their justified and true, if inexact, belief that men are vastly more likely than women to commit certain crimes, then this seems to me to qualify as profiling. It will be less efficient profiling the less precise and accurate their beliefs are about the actual ratios, but profiling nonetheless. Second, it is not obvious why we would want to draw a line between behavioural and other characteristics as Risse & Zeckhauser seem to suggest.[7] Thus, if customs agents give particular scrutiny to plane-passengers who purchased their ticket in cash, wear sunglasses while checking in and travel short-term (all behaviours) based on statistical evidence that such persons have a greater than average probability of being drug-smugglers,

---

[6] Risse and Zeckhauser's original formulation is that racial profiling consists in "any police-initiated action that relies on the race, ethnicity, or national origin and not merely on the behavior of an individual."

[7] Arguably, this is not the best reading of Risse & Zeckhauser's definition, which is intended to encompass only racial profiling in contrast to, perhaps, other forms of profiling. But if so, a more natural phrasing would be that racial profiling is any police action that relies on statistical evidence about correlations between race and crime *in addition to* evidence about correlations between other traits, behavioural and non-behavioural, and crime. This is close to the definition that I shall suggest below.

this seems as clear a case of profiling as the special scrutiny given to men above. What might mislead us here is the intuition that the kind of evidence obtained from immediately observable behaviour – from appearing nervous to brandishing a gun – is in some relevant sense qualitatively different from the kinds of evidence which involves inferences from general facts about the world, such as the correlation between purchasing a plane ticket in cash and smuggling drugs. This would be a mistake, because strictly speaking *no* evidence can avoid such inferences. Even taking nervous or threatening behaviour as evidence relies not only on inferences about what constitutes nervous or threatening behaviour – the relation between, e.g., facial expressions and body language on the one hand and states of mind on the other, and the notorious locus of intercultural misunderstandings – but also on the assumed (and presumably correct) correlation between such forms of behaviour and criminality. Nor can we avoid such deductive reasoning. We all necessarily and continuously form expectations about the traits other persons possess based on our observance of which traits they have demonstrated so far and our experiences of how such traits tend to correlate with others. In the words of Frederick Schauer: "It is simply how we think." (Schauer, 2003, p.75) It is not a crime to appear nervous, but police officers may have good reason to be more interested in persons who appear nervous than in those who do not if such behaviour

correlates with criminality.[8] And exactly the same reason obtains in the cases of "non-behavioural" and statistical evidence.

A parallel distinction might be thought to more successfully track our intuitions: that of the difference between reactive and proactive policing, and as an extension thereof, between allowing racial and ethnic characteristics to be used to apprehend criminals when based on evidence about a particular criminal and allowing its use based on statistical evidence about criminals in general. By "proactive policing" I understand any activity by the police which is initiated by the police, not as a response to knowledge that a specific crime has been committed, but on the suspicion that a crime *may* have been (or is in the progress of being) committed. "Reactive policing," by contrast, concerns situations where police respond to information that a crime has been committed.[9]

"Proactive police profiling," then, means the employment of statistical evidence in searching for and apprehending randomly selected offenders. Stop-searches, which are both the most

[8] It may not on such a broad a description. Probably, there are particular kinds of "appearing nervous" which warrant attention and others that do not. Nor am I suggesting that "appearing nervous" is likely to be sufficient in and of itself to raise probabilities to a level where subjecting such persons to a police practice is justified. I mean only that they can count in favour of doing so. I elaborate on these points in connection with the discussion of proxy sets and justifiability below.
[9] Here, as noted above, I limit myself to those police activities and practices whose purpose is the apprehension of criminals. Obviously, much reactive and in particular proactive policing will have different purposes, e.g., by striving to reduce crime-levels through other means than the apprehension of criminals.

controversial instances of profiling in public debates and at the core of the discussion in academic debates, are of this kind. "Reactive police profiling" means the employment of statistical generalizations in searching for and apprehending one specific offender. An obvious example of this is the kind of profiling famously employed by behavioural criminologists to identify and apprehend serial-killers.

More importantly, proactive and reactive policing may seem to be morally different, even in cases where they both involve giving special attention to persons because of their race. Thus, some would want to argue that doing so in reactive policing is not morally wrong in the way doing so in proactive policing is. As David Harris explains it: "It does make sense to use racial or ethnic characteristics in enforcement, but only in one context: *cases in which race or ethnic characteristics describe actual suspects.* […] In this situation, a description of the suspect's skin color serves not as a predictor of criminality, but as an identifying physical attribute that can be used, in conjunction with others, to determine whether a person observed might be someone wanted by the police as a suspect in an actual crime. […] What must *not* be allowed is using race or ethnic appearance, alone or in combination with other factors, to stop a particular person based on a *prediction* that he or she is more likely to be involved in crime." (Harris, 2002, p.152; cf. also Wasserman, 1996; Kennedy, 1997, p.137-138)

The problem with Harris' argument is that in both reactive and proactive policing it is exactly when and if a person's traits, such as race, serve as a predictor of criminality that the traits become

interesting to police officers. Consider how we might describe the epistemic situation of the agent who initiates the police-action. It is tempting to focus on what the evidence reveals about the offender, and to distinguish here between general evidence about groups of offenders and individualized evidence about a particular crime. But this is not the locus of the decision to subject persons to police-action. Whether given statistical or individualized evidence, police must first estimate how this affects the likelihood that an/the offender has a given trait (or set of traits), but second, and crucially, consider what the probability is that *this particular person* is an/the offender, *given* the likelihood now established that an/the offender possesses certain traits. While police-work thus necessarily involves both forms of inference, it is the later that is at stake in the decision whether to subject a person to a practice. If an eye-witness to a crime has described the offender as being of race R, then being of race R raises the probability that one is the offender, in just the same way that statistical evidence about, e.g., how gender correlates with crime does. This is what it means "to determine whether a person observed might be someone wanted by the police..." The relevant difference does not concern prediction. (cf. Harcourt, 2004, p.1342-1346)

A better explanation of the difference is perhaps the different proxies they involve. Thus: "It is not *racial* profiling for an officer to question, stop, search, arrest, or otherwise investigate a person because his race or ethnicity matches information about a perpetrator of a specific crime that the officer is investigating. That use of race – which usually occurs when there is a racially specific

description of the criminal – does not entail a global judgement about a racial or ethnic group as a whole." (Gross and Livingston, 2002, p.1415, my emphasis) This does not mean, however, that it involves *no* "global judgments about a group as a whole," or is not profiling. What it is, rather, is what we might call "testimonial-profiling," in that officers are moving from, e.g., eye-witness testimony via an inference involving the general validity of such testimony to the probability that someone who matches the testimony is the offender.

This is still an important difference, because different forms of profiling, i.e., the application of different proxy-trait correlations, will look very different morally. Some will perform better, that is, be easier to apply, more accurate, less vulnerable to bias, and some worse, that is, be more costly to use, less efficient, etc. We will return to this shortly, in the next section on costs and benefits.

A final consideration is worth introducing at this point: the difference between single proxies and proxy-sets. I have been concerned in the previous primarily with the use of single proxies, and in particular with the proxy of race, but in fact we normally never correlate between just one trait such as race and expected behaviour, but adjust for, e.g., gender, age, dress-code, facial expression, body-language and social context, even to form such superficial impressions as those available to police in stop-search practices. As Laurence Thomas observes: "With typical social monitoring, it is quite normal to mark the difference between a white male stranger in a tweed coat and a white male stranger in gang-like garb, and to suppose that an isolated encounter with the

former is less likely to be hostile than an isolated encounter with the latter. There is absolutely nothing about black males to suggest a different judgment is warranted between either a black male in a tweed coat and a black male in gang-like garb or, for that matter, a black male in a tweed coat and a white male in gang-like garb." (Thomas, 1992, p.32-33) In short, clothing-style – as well as quite a few other traits – is normally a much better proxy than race for predicting behaviour, although contrary to what Thomas suggests, race might still enter the picture *in combination* with any other relevant correlations in situations where race correlates with behaviours, e.g., crime.[10] When considering whether or not to subject a person to a practice, whether it be conducting a random stop-search or questioning suspects of a crime, police officers will apply not just single proxies but a proxy-set to estimate the likelihood that said person has committed a crime. Some proxy-traits may negatively correlate, making it *less* likely that the person has committed a crime, and balance other proxy-traits that positively correlate. Whether this leaves e.g. the nervous, old, poor, white woman or the calm, young, wealthy, black man as the statistically most probable offender will be a difficult question to

---

[10] Thus, Thomas concludes: "Regardless of the percentage of black hoodlums, it is rational to extend the statistic to all black youths only if black hoodlums cannot be reasonably and safely distinguished from other black youths, given a modicum of social monitoring skills and prevailing norms of self-presentational behavior." (Thomas, 1992, p.34; cf. also Cox, 1993, p.159-160) But, we might well insist, even in those situations where they can be distinguished in such a way a combination of clothing-style, race and any other relevant traits would, in the relevant circumstances, be a better set of proxies than clothing-style by itself.

answer, but it is not one police officers can avoid and the only way of dealing with it is to weigh all the relevant factors.

Thus, a proxy-set for a police-practice is a set of proxy-traits which jointly establish the probability of any given person being an identifiable offender,[11] so that for any given proxy-set:

> **Proxy Set S**. S is the set of proxy-traits (T, T_1, T_2...) that jointly establish probability P_S that a randomly selected person A with trait(s) T (and/or T_1, T_2...) will be identified as having committed a criminal act if subjected to police-practice PP.

Naturally, no proxy-set is better than the correlations on which it rests. These need to be non-spurious, and as accurately and precisely specified as possible. This is important, because there are literally an infinite number of traits that could be considered to be proxy-traits for criminality. The art of the police consists, as the quote which inspired the title of this article suggests, in ignoring

---

[11] Note that strictly speaking we need to distinguish sharply between the proxy-set which assigns probabilities that a person is an offender and the proxy-set which assigns probabilities that offenders will be identified by the practice in question. Although the two sets will likely contain mostly the same proxies they will assign different probabilities, given the inevitable imperfections of any practice in apprehending offenders subjected to it. In practice we are concerned primarily with the latter, and I shall restrict myself to this in the following, setting the differences between the two and the complexities introduced by distinguishing them aside.

the infinity of traits that have no bearing on the issue and picking out those traits that do.

Let me sum up the main points that I have attempted to establish in this section. Police use profiling in their various practices intended to apprehend offenders. By "profiling" we might initially understand the use of statistical evidence to better target offenders, but I have argued that there is no sharp dividing line between the way of targeting offenders which we associate with paradigm cases of profiling and other ways of targeting offenders; they simply apply different proxies in different ways. I have further argued that this also holds true across the division between reactive and proactive policing, that is, police actions that respond to known crimes that have been committed and those that attempt to detect or prevent unknown crimes, typically those in progress. Although it is tempting to describe these as morally different, and to allow otherwise impermissible characteristics, such as race, to enter police-work in the former case, the real difference concerns the proxy-traits at stake rather than the difference in kinds of police-work. Finally, I have suggested that profiling rarely, if ever, concerns only one proxy-trait. Instead, it will normally involve a set of traits which jointly provide police with (an estimate of) the probability that a person has committed a crime and can be apprehended if subjected to a given practice.

Thus, I would suggest, police profile for proxy trait T on practice PP *iff*:

1. Police believe T is part of proxy-set S (T, T1, T2 …), the members of which correlate with a relevant target-trait (criminality);
2. Police use S to target persons for PP, by basing their selection of which persons to subject to PP on the probabilities assigned by S.

Given this understanding of profiling as a broad and inevitable phenomenon, the question that I shall be concerned with in the remainder of this article is not "to profile or not to profile," but how.

The Benefits and Costs of Policing and Profiling

The strongest arguments for the use of racial profiling, it seems to me, must be consequentialist. Racial profiling, proponents could argue, help police to better target offenders when and if race correlates with crime, allowing police to catch more criminals. This is a beneficial consequence, so the argument will go, which justifies the imposition of an otherwise distasteful difference in treatment along racial lines. If not exactly common in the philosophical literature, this is certainly a line that is frequently suggested in public debates. However, as I hope to show in the following, the situation when viewed within a consequentialist framework is both considerably more complex and considerably less propitious for racial profiling. To demonstrate this I shall first explore the structure of costs and benefits in policing and the way that

profiling fits into this. Secondly, I shall outline three challenges that I believe racial profiling specifically faces.

This argument may be of interest, I hope, to non-consequentialists as well for two reasons. First, most forms of non-consequentialism will want to allow consequences to matter to some extent, even if the weighing of consequences is held to not constitute the full moral picture. Secondly, and perhaps more importantly, if as it seems to me the strongest arguments for racial profiling are consequentialist, and racial profiling fails on consequentialist terms, then there is little left to be said in its defence. If it cannot be defended on these grounds, then proponents will be very hard pressed indeed to argue their case.

Consider first the standard costs and benefits we might attribute to a police practice intended to apprehend criminals. These will vary, of course, depending on the ultimate value theory applied by the specific variant of consequentialism at stake, but most forms of consequentialism will converge on holding at least the following factors to have instrumental value, irrespective of the intrinsic value(s) which they ultimately serve:

Costs and benefits of criminal apprehending police practices

Resources	Deterrence
Inconvenience	
Punishment	Incapacitation

"Resources" here means the expenditure of these in terms of human-hours of police-work and the economic costs to society of training, equipping and paying police officers to invest these hours. These are resources that could otherwise have been invested differently by society in various beneficial ways. Allocating them to police practices thus counts as a cost of these forms of policing. By "inconvenience" I mean to cover those costs that accrue to the persons subjected to the police practice, in terms of the time they are required to set aside for it, the discomfort, pain or humiliation they may and will often suffer as well as any direct costs such as damaged clothing, vehicles or even personal injuries. Most persons, irrespective of issues of discrimination, will find, e.g., being stopped, questioned and physically searched by the police to involve such costs, even if only moderately so. The *Chicago Customs* case described in the introduction involves unusually severe inconvenience costs, but I do intend for this meek-sounding category to cover also the serious pain and humiliation that persons would likely suffer in that case. This cost, it is worth noting, applies to all persons subjected to the practice, both those found to be offenders and those who are not. Finally, "punishment" covers those costs imposed eventually by the criminal justice system on those persons found to be offenders by the practice. Given that they would not have had these costs imposed upon them had they not been subjected to the practice (at least, not on this occasion), the harm that criminal justice routinely

inflicts on convicted criminals in the form of e.g. incarceration must be counted as part of the cost of the practice.[12]

On the benefits side, the two benefits that I have included are both traditionally taken to be valuable because they contribute to decreasing crime, and thus preventing the harms caused by crimes. "Deterrence" is thus the prevention of crime by increasing the threat to potential offenders to a level where they are dissuaded from committing a crime that would otherwise have occurred. "Incapacitation" is the prevention of crime by specifically stopping a potential offender from carrying out a crime, and ought therefore to cover both crimes that are detected while in progress by the offender being subjected to the practice and any crimes that the offender would have committed had she not been subjected to the practice but cannot as a result of being identified, i.e., because she is convicted and incarcerated.

---

[12] Here, and elsewhere in much of the following, my analysis potentially faces the challenge of double-counting, i.e., if the punishment of criminals is held to be a cost of the police practice which apprehends them, it might be argued, it cannot at the same time count as a cost of other parts of the criminal justice system where we would traditionally like to include it, such as, e.g., the courts which hands out the sentence or the criminal law which mandates it. While I recognize this as an important issue, the limitations of this article preclude dealing with it and the underlying question of the relation between causality and cost-attribution. More importantly, as noted above, the sketch I am providing of the costs and benefits of police work serves first and foremost to illustrate the costs and benefits of profiling, and to set the stage for the challenges I shall present. These central points are not substantively affected by whatever view on the issue of double-counting one adopts. As such, I leave it open to the reader to discount the cost of punishment (and the benefits to which we turn below) by the requisite amount, if she feels that this is necessary to avoid double-counting.

Consider now what the costs and benefits of a given practice will be. Any realistic practice will only target a small minority of the population. Call these the positives, and the remainder the negatives. Further, for any realistic practice a group of the persons subjected to the practice will be innocent, indeed typically the majority will not be found to have committed a crime. Call these "false positives," and those identified as offenders true positives. Some of the costs and benefits will apply to false positives, while other costs and benefits will apply to true positives.[13] Notably, each instance of a false positive will incur full resource and inconvenience costs, but no punishment costs, and provide (presumably) partial deterrence but no incapacitation benefits. The benefit of partial deterrence can be presumed because even an instance of a police practice that fails to apprehend an offender is a public event which sends a signal to potential offenders that there is a risk of discovery and punishment. True positives on the other hand incur full costs and provide full benefits as per above. Finally, some costs and benefits will vary with the person upon which the practice is imposed. An ailing elderly person may suffer inconvenience more severely from being subjected to a police practice than a healthy adult subjected to the same practice. In the following I shall consider all costs and benefits to be averaged for

---

[13] The costs and benefits of negatives are incorporated in the account of costs and benefits of positives, in that the benefit of, e.g., deterrence consists in the reduction of a cost imposed by the total population. These need therefore not be counted separately.

the practice as a whole, so that each instance is considered to produce the same benefits and incur the same costs.[14]

Given the disparity between the costs and benefits provided by true and false positives, the ratio between them is clearly crucial to the costs and benefits of a police practice. Generally speaking, we assume that most, if not all, police practices rely on the positive cost-benefit ratio of their true positives to outweigh the negative cost-benefit ratio of their false positives. That is, although each instance of subjecting an innocent person to the police practice exacts a greater cost than the benefit it produces, the occasional instance of subjecting an offender to the practice produces a sufficiently great benefit, even considering its added punishment

---

[14] We need to carefully distinguish also between the costs that properly accrue to policing and profiling, and those that are often associated with but strictly speaking independent from policing and profiling per se. As Risse and Zeckhauser point out some of the wrongs often associated with racial profiling, such as the abuse of police powers to intimidate or harass civilians, are wrong independently of whether or not they occur in connection with profiling and even of whether or not they happen to be motivated by the same reasons which motivate profiling (such as racist beliefs and attitudes). What we are concerned with is whether and, if so, when racial profiling is wrong, which means whether it would also be wrong in situations where it did not coincide with other wrongs. Or, if we want to put it that way, whether racial profiling in and of itself constitutes a form of abuse by the police. (Risse and Zeckhauser, 2004, p.138-139; cf. also Lippert-Rasmussen, 2006b, p.XX; Lippert-Rasmussen, 2007b, p.XX) However, while in the following I set aside costs arising from such phenomena as racism, police brutality and the abuse of powers, I do think an argument can be made for including such abuses as have been specifically enabled by a practice or a form of profiling. That is, if a practice or form of profiling adds to the amount of abuse, say by increasing the contact between racist officers and minority persons that they would not otherwise have encountered, then the resulting abuses ought to have some weight as costs of the practice.

costs, that the practice in total is morally justifiable. This is the intuition driving the *Chicago Customs* case, where presumably the cost imposed on you by being searched as an innocent is much greater than any benefits produced, hence the only way to justify the practice as a whole is if there is a relatively high probability that people like you are offenders. It is not, however, a necessary truth that the cost-benefit ratio between true and false positives will take this form. We can imagine police practices, e.g., where even false positives have positive cost-benefit ratios, so that each time anybody, innocent or offender, is subjected to the practice, the total effect will be beneficial, or even practices where, perversely, false positives have positive cost-benefit ratios, but true positives have negative cost-benefit ratios, perhaps because of draconic criminal laws that lead to high partial deterrence benefits but extraordinary punishment costs, so that the practice relies on the instances of subjecting innocents to it to balance the instances of identifying offenders.

We can also, of course, imagine practices that fail to do more good than harm, if, e.g., both true and false positives have greater costs than benefits, or if the costs in one case are severe enough that the benefits of the other fails to outweigh them. Such scenarios need not be implausible.[15] Indeed, if the success-rate is as low as in a recent Danish case of stop-searches, where a 4 week intensive stop-search initiative directed against possessing and carrying

---

[15] See Jesper Ryberg's article in this issue of *The Journal of Ethics* for an argument that this is likely to be the case in realistic examples of racial profiling.

weapons resulted in 1 person being charged after 610 searches, we might well dismiss stop-searches irrespective of any issues raised by profiling. (Holst, 2009) Nor is there reason to think that this low rate of success is extraordinary. Over a 14 month period, Danish police performed 17.977 searches, finding 300 weapons, a hit rate of 1.6%. (Lindqvist, 2010) Similarly, in Amsterdam, the police spent 11.687 hours, conducted 32.332 searches and found 702 "weapons" (ca. 1 in 46) of which 15 were firearms (ca. 1 in 2155).[16] (Open Society Justice Initiative, 2009, p.note 238 & 239) I shall assume in the following, however, that we are discussing only practices that follow the first pattern identified, that is, practices where the positive cost-benefit ratio of true positives outweigh the negative cost-benefit ratio of false positives.

In any of these cost-benefit distributions, the ratio between true and false positives is provided by the probability of the practice successfully identifying a randomly selected person as an offender given the proxy-set used to select targets for the practice. Further, if we set as the minimal requirement for a practice that it must do

---

[16] The figures for the Netherlands as a whole are no less worrying: "A 2005 Dutch study of the efficiency of preventive searches for weapons in eight cities over a two-year period found that the searches disproportionately targeted minorities and that the hit rate was only 2.5 percent [...] this figure was inflated by the inclusion of items such as penknives." (Open Society Justice Initiative, 2009, p.53)

more good than harm,[17] we can summarize the costs, benefits and justification in the following terms:

Costs, benefits and minimal justification of police practices:

1. Police practice PP has probability P_1 of identifying a randomly selected person A as having committed a criminal act, i.e. $P_1 = P(A_{true} \mid PP)$;

2. True positives produce benefits B_{true} and costs C_{true};

3. False positives produce benefits B_{false} and costs C_{false};

PP is *minimally justified iff*: The expected benefits of PP outweigh the costs, i.e. $(P_1 \cdot B_{true} + P_1 \cdot C_{true}) + ((1-P_1) \cdot B_{false} + (1-P_1) \cdot C_{false}) \geq 0$.

The role of profiling in this should be clear. Since the ratio of true positives to false positives is determined by the probability of PP identifying offenders, and since this probability is a function of the proxy-set applied to target offenders, the profiling employed is crucial to the justifiability of a practice. The potential benefits of

---

[17] I shall return to what it means for a practice to be maximally justified, i.e., to meet the consequentialist requirement of being not only beneficial, that is to have better consequences than inaction, but optimal, that is at least as good as any available alternative action. Cf. the section on the valuation challenge, below.

profiling are therefore reasonably obvious and uncontroversial: by applying non-spurious proxies to refine selection one will increase the benefits derived from apprehending an offender or decrease the costs in terms of police resources and inconvenience.[18] Consider an example:

> Profiling 101. Police practice PP_1 subjects 101 persons to the practice every day. These 101 persons are selected from a much larger group of 10.100 potential subjects. Further, in each instance of subjecting a person to the practice police select the person from a much smaller group of presently available subjects, say 20, who they, based on proxy-set S_1, consider to be the most likely offender. These have, as it turns out, an average probability of 0.05 of being an offender, meaning that roughly five offenders are caught per day, while 96 other subjects are

---

[18] Frequently, the first type of benefit, the increased apprehension of offenders, is granted the most attention. But the second type is important because it means that profiling may be justified, even in situations where apprehending offenders is not desirable. Suppose for instance that the effects on society-wide crime levels from apprehending more offenders are negligible, but that the law imposes draconic punishments on apprehended offenders. In that situation, it may not be desirable to increase the number of apprehended offenders, and so profiling cannot be justified in this respect. But it may still be the case that it increases police efficiency, so that it would be possible to apprehend the same number of offenders with fewer police resources, freeing those resources for other potentially beneficial types of police work, or even for altogether different policy initiatives.

false positives. Now we add non-spurious proxy T to S_1, but otherwise leave the practice unchanged. S_{1+T} will allow police to be somewhat better, depending on the strength of T, at selecting the one person among the 20 with the greatest probability of being an offender. Those persons in the 20 who are T-persons will have their probabilities increased somewhat, to the extent that if the most probable offender on the basis of S_1 was a non-T-person, then a different person may now be the most probable offender on the basis of S_{1+T}. The difference between profiling and not profiling for T is the number of times that this happens. Every one of these actions has a slightly higher probability of apprehending an offender than the practice did just using S_1. As it turns out, the persons selected by S_{1+T} have an average probability of 0.075 of being offenders. This means, if the number of actions remains unchanged, ca. 2.5 more true positives per day presumably increasing overall benefits, or, if the number of apprehensions is kept constant, a decrease of ca. 33 actions daily presumably decreasing overall costs.

Profiling, in short, increases the efficiency of a practice by better targeting, decreasing the number of false positives per action and increasing the number of true positives per action.

Note the importance of proxy-sets here. If we disregard the existing set when adding a proxy-trait, or if we imagine a situation where the set has only one member, then we would get a situation where any person with the trait is more likely to be an offender than any person without the trait: "Say eyewitness testimony suggests that there is a 60 percent chance that a crime was committed by an African American man, and African American males make up 25 percent of the population; one should then inspect *only* African American males, and *mutatis mutandis* for other scenarios. The reason is that an African American male is 2.4 = 60%/25% times as likely to be guilty as a person selected at random."[19] (Risse and Zeckhauser, 2004, p.140) Although Risse & Zeckhauser point to several qualifications for this view, it still

---

[19] Our intuitions might steer us wrong here: if it is the case, e.g., that non-T-persons have 0.01 probability of being offenders and T-persons have a 0.04 probability then it might seem that an appropriate rate of stop and search is 1:4. But this is not true: in every single case where the police can decide whether to stop either an T-person or a non-T-person they are choosing sub-optimally if they choose to stop a non-T-person. An easily graspable comparison is with lottery-tickets, where we would not want to buy tickets with a 1% chance of winning if there are other tickets with a 4% chance of winning. Assuming that the tickets are identical in other respects, such as price and prize-size, the obvious choice is not to buy tickets in a ratio of 4:1, but to buy as many tickets with a 4% chance of winning as possible and to only start buying the 1% tickets once all of the 4% tickets have been sold (although strictly speaking, it is never rational to buy lottery tickets, because no lottery offers a better than even return for money).

seems to me to inaccurately characterise the demands of efficiency, at least in any realistic scenario where police will have more proxies to go by than gender and race. Rather, the practice of inspecting will involve certain costs and benefits. For some P_1 the expected costs and benefits of subjecting a person to inspection will balance. Call this threshold-probability N. Whenever police have the opportunity of subjecting a person with $P_S > N$ to PP the expected consequences are therefore beneficial, and as above, I shall call it minimally justified for a police agent to subject a person to police practice PP when that person's probability of being an offender is equal to or greater than N. Conversely, police ought to refrain from subjecting persons with $P_S < N$ to PP; in these cases the expected costs will outweigh benefits. In summary:

> **Minimally justified police action.** It is *minimally justified* for a police agent to subject person A to police practice PP *iff* $P_S \geq N$, where $N = P_1$ such that $P_1 \cdot B_{true} + (1 - P_1) \cdot B_{false} = P_1 \cdot C_{true} + (1 - P_1) \cdot C_{false}$.

Consider how this will affect police practices, in a variation of the example above:

> **Profiling 101a.** Suppose that a neighbouring police department still using proxy-set S_1 reviews police practice PP_1, and decides that N = 0.1. The existing practice thus does more harm than good and would continue to do so even with the refined selection of S_{1+T} — the bar

needs to be set higher. As a result, police continue to apply S_1 but only subject persons where $P_S \geq 0.1$ to PP_1. Suppose also that, unlike above, they never subject only the most probable in a group, but *all* persons encountered where $P_S \geq 0.1$. Their selection is still substantially narrower than before; they now subject only 40 persons daily on average, as these are the number of suitable targets locatable with same expenditure of resources.[20] Suppose that the resulting average probability is 0.125, resulting in roughly the same number of apprehensions (five daily), but a much lower number of false positives. The latter are the reason that PP is now overall beneficial. At this point T is introduced, leading to S_{1+T} being used to select targets. What happens? Much as above, some T-persons will see their probabilities shift above 0.1, while some non-T-persons may see theirs shift downwards below 0.1 (assuming negative correlation). The result? Some T-persons who ought to be subjected to PP, but

---

[20] Note that the police department ought to increase resources until all persons with $P_S > N$ are subjected to PP, but doing so complicates matters because increasing costs in this way will raise N. There will be some point of optimal efficiency balancing these two conditions, but while finding it is a point worth pursuing in practice it is not relevant to the argument I am pursuing here.

would not have been under S_1, will be targeted using S_{1+T}, while some non-T-persons who ought not to be subjected to PP, but would have been using S_1, will not be targeted under S_{1+T}. As such the composition of the group targeted will change, and average probability of apprehension for PP will increase ($P_1 < P_{1+T}$) with accompanying benefits.

Returning to Risse & Zeckhauser's example, if police use a proxy-set, rather than the single proxy provided by the evidence described, there may be whites and women with higher probabilities of being the offender than some black men, just as there may be blacks and men who do not have a sufficiently high probability of being the offender, in spite of the strong correlation established by the eye-witness testimony, to be legitimate targets for inspection. Imagine for instance that police know of one young, white woman who is a habitual committer of the crime in question, lives in the vicinity of the crime and has an appearance that could be mistaken for a young black man in conditions of confusion and stress, as well as of an elderly, law-abiding black man who was plausibly not in the vicinity of the crime at the time in question. Is it unrealistic to suppose that the first might have a high enough probability of being the offender to be worth inspecting, while the second does not? Possession of a trait that correlates with criminality makes it more likely that one's probability of being an offender has crossed the threshold beyond

which police ought to subject one to a practice, but it does not in and of itself guarantee it, no matter how strong the correlation.

What then of the costs of profiling? The costs most often associated with profiling concern its unintended side-effects, which I think can roughly speaking be said to fall in 4 groups: anti-deterrence, alienation, stigmatization and promoting structural inequality. Let me briefly sketch each of these in turn.

Anti-deterrent effects emerge from the fact that focusing police resources on one group will decrease the amount of resources available for actions against other groups. In the most extreme case, a police policy that targeted *exclusively* a specific social group which had been determined to be the most likely offenders for a type of crime would in effect remove the deterrent effect upon all persons not members of this social group, by guaranteeing them that they would not be subject to police scrutiny. (cf. Risse and Zeckhauser, 2004, p.141; McGary, 1996) In reality this cost is compounded, as Bernard Harcourt shows, by the way different groups may respond differently to marginal changes in the deterrent threat. To briefly restate his point, we can imagine a city with a deprived minority who are more prone to criminality but also highly resistant to variations in deterrence-incentives, perhaps because they lack adequate alternatives to the criminal career. What this means is that increased profiling of this group, though more efficient in terms of successfully targeting offenders, will produce a disproportionately small deterrent effect, easily offset by the rise in offences among the majority-group resulting from the decreased deterrence they experience. (Harcourt, 2004, p.1296-1303)

Alienation effects stem from the fact that, irrespective of whether profiling is rational or not, there will be situations where those profiled experience the increased attention as an expression of distrust, disrespect and/or discrimination. This can have a number of implications, from less cooperation with the police force, to increased criminality because the legal system, its norms and its enforcers are no longer viewed as legitimate. At the extreme end, one frequently cited problem with profiling young Muslim men for terrorism is the risk that this will lead some of them to sympathise with the terrorism that the profiling is intended to prevent. While it is unlikely that profiling will in and of itself make a person support, much less actively partake in terrorism, the perceived injustice of being profiled could in some cases be the nudge that pushes a wavering potential criminal, terrorist or otherwise, to going through with the deed.

Stigmatization effects stem from the impression of the profiled group created by profiling, both on those persons subjected to profiling and on others (police officers, non-profiled members of the group, non-members of the group). While a profiling practice ought to rely on rational attitudes regarding the profiled persons, i.e., by applying statistically correct correlations, it is both possible and likely that profiling can engender irrational attitudes about the target group. Thus, as Arthur Applbaum notes, it is all too easy to slip from a rational correlation to over-generalizing, as well as to associating correlations with blameworthiness, and so for police officers: "As the proportion of true positives picked out by a strategy rises, the cognitive discipline required to maintain

respectful treatment in group-based patrol is enormous." (Applbaum, 1996, p.155) But this applies equally, although in different ways, to the ways members of the profiled group will perceive themselves, and the way the rest of society will perceive them.

Finally, *inequality* effects occur for several reasons, first and foremost of which is that social deprivation is certain to follow increased criminal supervision, both directly in the shape of incarceration, and indirectly in the shape of lessened job prospects and the lower income that this causes, damaged family structures and individual brutalization. But also as a result of the way both profiled persons and non-profiled persons may respond rationally by seeking segregation. If my chances of being subjected to a stop-search increase when I give my colleague a lift, because she possesses trait T, then it makes sense for me to give a lift instead to a different colleague who, like myself, does not. Inequality is thus likely to increase as profiling concentrates burdens that would otherwise have been more widely distributed on a particular group. Concentration is also likely to create added burdens, such as those that follow from rational strategies of segregation, and from cumulative effects of deprivation, since the groups targeted are likely to be groups that are already socio-economically deprived given how strongly socio-economic deprivation normally correlates with the target trait of criminality.

Overall I share Annabelle Lever's concern that: "…we should expect racial profiling to exacerbate racism in society at large, even in apparently unrelated areas such as housing, transport,

employment, and entertainment. […] It is likely that racial profiling discourages black people from living, travelling and working in white neighborhoods, especially at night, and so compounds residential and occupational segregation. It is likely that it discourages black people from joining the police, and so perpetuates a damaging public perception of the police as hostile to black people. It is likely that racial profiling obscures the fact that most violence is intra-racial, rather than inter-racial and committed by a minority of people, whatever their color." (Lever, 2007, p.23; cf. also Loury, 2005; Harris, 2002; Harcourt, 2004, p.1329-1330)

Two factors which will heavily influence these costs are the publicity of the profiling and the social character of the group profiled. The publicity of profiling is likely to have a strong impact on deterrence, alienation and stigmatization, whereas the social status of the group is most likely to influence alienation, stigmatization and inequality promotion. Generally speaking, publicity, that is, the recognition that profiling is being applied, will increase the redistribution of deterrence, aggravate alienation and deepen stigmatization. On the other hand, if hypothetically a form of profiling could somehow be implemented without anybody realizing that it was in use, it would be unlikely to incur these costs.

Simultaneously, the social character of the group profiled will play an important role. Most of the concerns regarding profiling focus on its use to target groups that are both highly socially salient, centrally racial, ethnic and religious groups where "…perceived membership of it is important to the structure of social

interactions across a wide range of social contexts" (Lippert-Rasmussen, 2007b, p.386), and which are socio-economically deprived and/or marginalized. Obviously, targeting some groups will have a greater impact than targeting others, say, "blacks" versus "bankers" (for financial crimes), both because the identification of the group *as* a group is stronger, and because bankers are not commonly perceived nor do they perceive themselves to be in a vulnerable position or the victims of historic injustice at the hands of the state. Or, as David Wasserman aptly describes how most persons might experience it: "It is far less demeaning to incur suspicion because of one's accidental or voluntary associations than because of one's race. It is, for example, merely annoying to become an object of close police scrutiny because one attended an opera at which gunshots were fired. It is profoundly demeaning to be subject to close police scrutiny because one is a black male youth." (Wasserman, 1996, p.117) These qualifications make it even more difficult to generalize about the quantity of costs attached to different types of profiling. Such issues must be settled on a case-by-case basis, taking the social context into consideration.

In conclusion, we should be aware that even in situations where the beliefs upon which profiling are based are true it is unlikely that the attitudes it will promote are rational. It is far more likely that applying a form of profiling, particularly one involving highly socially salient proxy-traits, no matter their accuracy, will also promote unfounded biases. Secondly, we should expect profiling to have unintended bad consequences, such as the segregation of

residential areas and the loss of deterrence, which are neither emotional nor personal, but reflect rational strategies for responding to the situation of profiling. Such costs must be included when weighing the pros and cons of a profiling practice.

To sum up, profiling involves the following benefits and costs:

Benefits and costs of profiling

Increased benefits through higher ratio of true positives	Anti-deterrent effects
	Alienation effects
Decreased costs through lower ratio of false positives	Stigmatization effects
	Inequality promoting effects

Furthermore, building on the definitions previously established, we may now define the conditions of "minimally justified profiling:"

> Costs, benefits and minimal justification of profiling:
> 1. Police practice PP using proxy-set S has probability P_1 of identifying a randomly selected person A with trait(s) T_1 (and T_2, T_3 ...) as having committed a criminal act;

247

2. Police practice PP using proxy-set S_{+T} has probability P_{1+T} of identifying a randomly selected person A with trait(s) T (and T_1, T_2, T_3 ...) as having committed a criminal act ;

3. True positives produce benefits B_{true} and costs C_{true};

4. False positives produce benefits B_{false} and costs C_{false};

5. Using S_{+T} instead of S adds costs (anti-deterrence, alienation, stigmatization, inequality) C_{+T}.

Profiling for trait T on police-practice PP is *minimally justified iff*: The marginal benefits of using S_{+T} outweigh the costs, i.e. $((P_{1+T} - P_1) \cdot B_{true} + ((P_{1+T} - P_1) \cdot C_{true}) + ((1 - P_{1+T}) - (1 - P_1)) \cdot B_{false} + ((1 - P_{1+T}) - (1 - P_1)) \cdot C_{false}) \geq C_{+T}$

While hardly innovative in terms of normative theory, nor indeed particularly refined in terms of probabilistic thinking, I believe that this captures the essentials in a concise format. And it explains our conflicting intuitions, as I attempted to illustrate them initially by the opposed pull of *Chicago Customs* and *London Metropolitan*. We want at once to focus the efforts of the police only on those who deserve it, because this minimizes the costs imposed on the innocent, and to not focus the efforts of the police in a way that incurs costs by alienating, stigmatizing and promoting inequality.

Weighing these concerns is a central challenge in the moral assessment of police-work. Having thus set out the framework for assessing profiling in consequentialist terms, let me turn at last to the reasons why I think racial profiling is unlikely to be justifiable.

Three Challenges – Foundation, Valuation and Application

In the following I will sketch out three challenges that racial profiling faces. Together, I believe they constitute at least a prima facie case against racial profiling, as well as illustrations of why proponents will have a hard time meeting this prima facie challenge. The first concerns the relation of racial profiling to potentially unjustified police practices, as the glossing over of unsound foundations, the second concerns the uncertainties of the costs and benefits of racial profiling, but particularly the benefits, and the third concerns the difficulty of properly applying racial profiling and the diminished efficiency likely to arise from cognitive and psychological biases.

The Foundation Challenge

Given that consequentialism is committed to holding an action to be morally wrong unless it is at least as good as all available alternatives, i.e., the maximisation of good consequences, the requirement of being minimally justified may seem to fall trivially short of the goal. As Applbaum observes: "Among the set of search strategies with positive net benefits, some are better than others. There may be a more refined strategy that is more

efficient... [...] If a more refined search strategy is available, not to use it is inefficient."[21] (Applbaum, 1996, p.146)

Consider, however, under what circumstances it is true that there exists a potentially superior alternative. What are the given alternatives to profiling for T on PP? First and foremost, it is clear that profiling in the light of all relevant proxies is not only difficult, given the high uncertainty of many factors, but also requires resources in terms of time and effort, resources which will frequently not be available in the relevant situation, or which might be better spent in other ways in the relevant situation. A police-officer driving past a person on the street and considering whether to stop and search her does not have time to make the kind of complicated probabilistic or moral arithmetic outlined in this article. As Stephen Maitzen argues: "...a social policy involving a given level of [statistical discrimination] is justified if and only if the information-cost of further statistical refinement equals or exceeds the net social utility to be gained by such refinement." (Maitzen, 1991, p.26) Requiring police to use maximally efficient profiling must therefore be understood as the maximum effort compatible with optimizing the ratio of profiling benefits over

---

[21] Cf. also e.g. Lippert-Rasmussen who suggests that "unalloyed racial profiling" relies, among other criteria, on the fact that: "...no alternative, and equally or more effective, way of doing this is feasible." (Lippert-Rasmussen, 2006b, p.191-192) and Howard McGary, who argues that targeting searches against members of one particular race, for the purpose of apprehending drug couriers, is not justified even if the correlations is non-spurious unless: "...there is a high probability that stopping only speeders from this group would maximize arrests for drug offenses". (McGary, 1996, p.137)

selection costs. Secondly, an action need only be considered an alternative if doing it and profiling for T on PP are mutually exclusive. If they can be performed simultaneously and will both be beneficial then consequentialism will require us to do both, but even if one is clearly superior to the other this could not lead us to be morally required to perform the superior *rather than* the inferior action.

Keeping these considerations in mind, maximising profiling looks relatively simple because the only obvious alternative to profiling for T on PP is PP without profiling for T. It may be possible, though I am hard pressed to think of an example, that profiling for race, i.e., including race in the proxy-set which is applied to select targets for a practice, could preclude other forms of profiling, but it does not seem likely. Thus "a more refined search strategy" than racial profiling, if racial profiling is independently beneficial, will be racial profiling plus whatever other proxies refine the search strategy, rather than such other proxies without racial profiling. In this case it seems that the relevant alternative to racial profiling is not to profile for race, which means that minimally justified profiling *is* maximising.

This does not, however, get police practices off the hook. The obvious alternative to a police practice is also no police practice, i.e., the same situation without police subjecting anyone to the given practice. This is the alternative covered by the requirement of minimal justification. But the stricter requirement of consequentialism forces us to also explore whether there are other and more efficient ways of promoting the values which ultimately

251

justify police practices, such as urban renewal, improved educational opportunities, etc. Again, unless doing one prevents us doing the other, these do not count as alternatives in the pertinent sense, and this may not be true of many superior social policies aimed at the same purposes. But any police practice will also have a near-infinite number of variations where details in the practice are altered, which really are incompatible with the practice: we cannot ask officers to apply two different practices to the same situation simultaneously. Thus, when assessing practices we really do need to think carefully about maximising, not just minimal justification. Is this not a problem, however, for assessing practices rather than racial profiling? Perhaps not exclusively. For racial profiling, when applied to a practice that does not meet the requirement of maximisation, and in some cases not even minimal justification, risks supporting its continuation and prolonging a practice that is morally wrong.

As I attempted to show in *Chicago Customs* initially we can easily imagine practices where the costs imposed seem out of proportion to the benefits potentially produced. The answer in such situations could in one sense seem to be more profiling, rather than less, since a practice that imposes heavy costs can only hope to achieve minimal justification by having a high probability of apprehending offenders. But in realistic scenarios we might worry that something else will happen instead. The fact that racial profiling tends to target marginalized minorities may make it easier to ignore the fact that a practice is not, in fact, justified. In the case on which *Chicago Customs* is based this seems to have been one of the effects. It was

clear to those subjected to the practice that there was no way it could be justified, but at least partly because they belonged predominantly to the twice marginalized group of black women their complaints fell on deaf ears for far too long. It took the lucky break of one subject encountering an interested reporter, and the resulting massive media attention, for authorities to take seriously the plight it had imposed on innocent travellers and reassess the practice. (Harris, 2002)

Because racial profiling tends to redistribute burdens so that the costs of a practice are imposed on those least capable of protesting or refusing them, it risks supporting practices that – profiling or no – are not minimally justified. And because such police practices represent genuine social evils this risk must be counted against it. Polemically put, if racial profiling is built on the morally corrupt foundations of unjustified practices, and protects these from the rl>http://www.archive.org/stream/lelysrouge00franuoft</url></related-urls></urls></record></Cite></EndNote> (France,

The Valuation Challenge

An important issue that I have thus far neglected to address concerns how exactly to weigh the various costs and benefits. Clearly, we are dealing with costs and benefits that are not immediately commensurable. Even supposing that we had accurate figures for the costs and benefits involved, how ought we to weigh a certain degree of stigmatization of a racial group and the expenditure of so-and-so much money and time against the prevention of a specific number of crimes? This is clearly a case of

apples and oranges. Nor am I suggesting that there is – or at least that I know of – any easily applicable method of translating such disparate goods and evils into a common currency. First, because I have deliberately attempted to shape my analysis in a way that did not rely on any one particular value-theory I do not have resort to the intrinsic value(s) that these instrumental values serve, and which would be needed to assess commensurability. Second, and more profoundly, because even if we were to adopt a well-developed value-theory, e.g., hedonism with QALYs as the measuring unit, I do not want to claim that there is a simple and uncontroversial way of exchanging the different costs and benefits. Most of the work here will, I suspect, have to be done by intuitions, despite the serious methodological concerns we ought to have about these. (cf. e.g. Singer, 2005) However, this is a problem which affects costs and benefits equally, and in a wider context a challenge for any ethical theory which grants consequences any weight. As such it does not constitute an argument for or against racial profiling in itself, but one difficulty we face when we want to assess it.

A different, but related problem is that the empirical facts of the matter are not easy to sort out. Obtaining reliable statistics about racial differences in the propensity for crime has turned out to be extraordinarily difficult, as has obtaining reliable statistics for the deterrent effect of police practices or criminal justice in general. (cf. e.g. Von Hirsch, Bottoms et al., 1999; Harcourt, 2004, p.1358-1371) But the difficulties involved here, and the resulting uncertainties about the size of the instrumental benefits and costs,

254

do not affect benefits and costs equally. On the benefits side, incapacitation is both measurable and predictable. We can tell exactly the incapacitory effect of a given practice and anticipate with relative confidence the effects of increased (or decreased) apprehensions resulting from adding or removing a form of profiling. Similarly, we can make relatively good assessments of any efficiency benefits in terms of resources and inconvenience. Both are easy to observe and the changes resulting from increased or decreased efficiency can be measured and anticipated. But not so for arguably the most important benefit: deterrence. This is both uncertain and difficult to measure. Even assuming that racial profiling increases the efficiency of a practice, there is no guarantee that this marginal increase will lead to greater deterrence or easy way of finding out how much of a difference it makes if it does.

On the costs side of the equation effects are more certain. The marginal increase in punishment is as easy to observe and anticipate as the use of resources and incapacitation effects. Further, there can be little doubt that conditions will, in those cases where the correlation between race and crime justifies racial profiling, also make race a social phenomenon subject to the costs I have described. Some members of a profiled racial minority will almost inevitably feel ill-treated and respond with hostility towards the legal system, just as both some members of the minority and the majority will almost inevitably internalize negative stereotypes about the minority, or reinforce existing prejudices. And since profiling redistributes costs many of the inequality effects will occur simply as unintentional or rational responses to changes in

the incentive structure. All these costs may be difficult to measure precisely, but there is little doubt that they will occur. Only anti-deterrent effects are subject to the same severe uncertainty that afflicts deterrence, and for the same reasons. Harcourt, however, summarizes his review of the data in that: "Given the paucity of the evidence on both relative elasticities and offending, the conclusion is tentative, but under these assumptions [relatively low general elasticity, slightly lower elasticity for minorities and slightly higher natural minority offending rates], racial profiling probably increases the profiled crime." (Harcourt, 2004, p.1371-1372)

Uncertainties, it seems, do not afflict both sides of the equation equally. Racial profiling is a form of profiling that carries heavy and plausible costs. Other forms of profiling will suffer from similar uncertainties on the benefits-side, but not need to weigh these against such heavy burdens of justification on the costs-side. With the benefits so uncertain, it may be hard to achieve minimal justification, a problem which is aggravated by the difficulty of maintaining the efficiency of racial profiling in its application.

The Application Challenge

The third and perhaps most serious challenge for racial profiling is that there is good reason to be sceptical about the possibilities of properly applying it. The challenge here can be traced to two factors: the difficulty of obtaining accurate and precise data about the correlation between race and crime mentioned above, and the further difficulty of using this data in an efficient way because of cognitive biases, which is the topic here. Suppose for an instant

that reasonably accurate and precise data could be obtained, what is the likelihood that it could be effectively applied? Not, I would suggest, as good as we would prefer. This is so because, as studies in decision theory and behavioural economics has demonstrated in increasingly convincing detail, we humans are not very good at coping with probabilities.

Of the many problems afflicting the application of probabilities in drawing inferences, one of the most serious is undoubtedly the failure to appreciate the influence of the base rate. Consider the following example, lightly modified from how it is presented in (Schauer, 2003, p.94-96):

> **Schauerville**. A violent crime has been committed in a city where 85% of the population is white and 15% of the population is black. Crime-rates for the two races are known to be equal, but an eye-witness to this particular crime, e.g. the victim, claims that the assailant was black, a piece of evidence the reliability of which is estimated as 80%. Is it likely that the offender is black?

Our intuitions might lead us to combine the high probability of the witness being correct with the fact that since blacks make up such a small proportion of the population, and the witness identified the assailant as a black person, then it really must be so. But the answer is a counter-intuitive no. There is, in fact, a roughly 0.59 probability that the assailant was white. The assailant is white on

257

average in 85 out of 100 cases and the observation only partially redistributes this base probability, because a much higher number of white offenders will be misidentified as black than vice versa.[22]

In addition to the problem of insensitivity to the base rate at least two further distortions may importantly taint probability estimates. Since race is such a highly salient trait our cognitive biases make it highly likely that its predictive power will both be over-estimated, and that it will serve as the starting point of prediction after which an "anchoring" effect will take place, reducing the predictive power, and hence adjustment of the estimated probability, of other traits. Correcting for these biases is difficult, because the calculations are complex and the results counter-intuitive. Suppose that we expand the example to further illustrate the difficulties police will face when they try to use this evidence in practice:

---

[22] By Bayesian inference P(white): 0.2 · 0.85 / (0.2 · 0.85 + 0.8 · 0.15) = 0.5862 Note also the way that unreliability will lead to exaggerated perceptions of the correlation of minorities with a trait. Suppose Schauerville is representative of average witness reliability in reporting offender race, i.e. the city has an 85% white and 15% black population, there is no correlation between race and criminality and witnesses are, on average, 80% reliable when reporting the race of the offender. Assume further, and probably unrealistically, that the reliability is the same no matter the race of the offender, i.e. that white offenders are as likely to be misidentified as are black offenders. The resulting *observed* crime rates work out at a shocking 29% crimes committed by blacks, and 71% committed by whites, which deceptively indicates that far from being equally criminal Blacks are roughly twice as criminal as whites. And this warping of the picture emerges, in this example at least, without any form of racial prejudice, simply because many more whites will be misidentified given that there are more of them. Put differently, there is a "regression towards the mean"-type effect in any situation in which populations are uneven and trait attribution unreliable.

Schauerville II. Police are now looking for the suspect, and considering whether to stop and search person A, a practice for which they estimate the probability of that person's being an offender necessary to obtain minimal justification (i.e. $P_I \geq N$) as 0.02. Their proxy-set allows them to consider the following traits:

1. A is black; blacks are 15% of the population; evidence gives a 41% chance that the offender is black;

2. A is male; males are 50% of the population; 90% of violent crimes are committed by men;

3. A is young, youth are 20% of the population; 50% of violent crimes are committed by youth;

4. A looks nervous; 5% of persons observed by police look nervous; 50% of offenders observed by police look nervous;

5. A is in the vicinity of the crime-scene; 100.000 people are in this part of the city; there is one offender among them.

Ought police subject A to the practice? Obviously, even if they were lucky enough to somehow have such exact figures, police officers cannot be expected to work out mathematically the probability of A being the offender, but will have to rely on their

intuitive estimate of how likely the traits combine to make it. Given the low requirements – police can afford to search 49 innocents for every offender – and the combination of relatively many suspicion-raising factors it may seem that the answer is yes. In fact, A's probability of being the offender is only about 0.0012, well below the 0.02 threshold where subjecting him to the search could be expected to do more good than harm.[23] Importantly, none of the problems I've sketched here represent flaws on behalf of individual agents, in the sense of psychological distortions such as racial biases and the like. As Daniel Kahneman and Amos Tversky put it: "These biases are not attributable to motivational effects such as wishful thinking…" but rather represent commonly used cognitive heuristics that while "highly economical and usually effective" also lead to "systematic and predictable errors". (Kahneman and Tversky, 1974)

We must not ignore the influence of psychological distortions, though. Added to the systemic weaknesses of our thinking about probabilities described above is the spectre of racism which, much as we would prefer the situation to have improved, is likely to at least occasionally haunt even modern police departments. The problem may not be as bad everywhere as e.g. suggested by the OSJI-report, which holds that the opinion expressed by a Spanish

---

[23] Start by assigning A the 1:100.000 probability of being the offender based on the number of people in the vicinity of the crime scene. We can then adjust this by Bayes' theorem for being black $((0.00001 \cdot 0.41)/0.15)$: 0.0000273, being male $((0.0000273 \cdot 0.9)/0.5)$: 0.0000492, being young $((0.0000492 \cdot 0.5)/0.2)$: 0.000123 and looking nervous $((0.000123 \cdot 0.5)/0.05)$: 0.00123.

police officer that "all murders are related to immigrants (as are) 90 percent of drug crimes and gender violence" reflect commonplace perceptions in Europe. (Quoted in: Open Society Justice Initiative, 2009, p.35) But the question is still how widespread such prejudices are, not whether they exist. The existence of police prejudices does not in and of itself affect the justifiability of profiling. Like police abuse it constitutes a problem independently of profiling and the obvious solution is to work at eradicating such prejudices, whether or not profiling is being applied. But as long as it exists institutional racism is worrying also because it deepens our concerns over the efficiency of profiling. Racist police officers are likely to further over-estimate the importance of racial differences in crime-rates, weakening the beneficial effects of using even a non-spurious correlation between race and crime.

Consider how systematic over-estimation of the effect of race as a proxy-trait would affect profiling. As illustrated in Profiling 101a the marginal benefits of profiling is the difference it makes in the average probability of a practice apprehending offenders. It does so by helping police estimate the individual probabilities of persons and target only those persons where $P_i \geq N$ and, in situations where they must choose, the person with the greatest probability of being an offender. Overestimating the effect of race will lead to police subjecting some persons of the profiled race whose probability is actually lower than N to the practice, and it will lead them to sometimes prefer persons of the profiled race to persons of a different race when the latter would accurately have been estimated to have a higher probability of being offenders. In

both cases this will decrease the resulting frequency of true positives and increase the frequency of false positives.

How bad the results of overestimation are for profiling depends on two factors. The first is the degree of overestimation. The second is the predictive power of race as a proxy-trait for crime; the higher this is the greater the margin of error. If the predictive power is sufficiently strong and misestimation is sufficiently small then profiling will maintain its benefits. But as the first grows weaker or the second greater, the benefits of profiling will diminish. Eventually, if the two factors cross a certain threshold, profiling will *reduce* the benefits of the practice in addition to still carrying its full costs.

Although the degree of inefficiency inherent to the application of a given proxy will, of course, depend on the specific situation in which the proxy is applied, it seems to me that there is good reason to be sceptical that proxies that, like race, are at once costly to use and have relatively limited predictive power will sustain enough of their marginal benefit to meet the minimal justification.

Conclusion – the Art of the Unseen

Throughout the course of this article I have sketched both what I take to be the consequentialist framework for assessing racial profiling, by separating profiling from the practices in which it is applied, reviewing the costs and benefits of each and establishing the conditions of minimally justified profiling. Doing so enables us to tackle the best arguments for racial profiling head on, as these

tend to rely exactly on the proposed benefits of targeting racial groups when and where membership of such a group correlates with crime. However, I have argued that the situation is considerably more complex than is sometimes assumed, and in particular that racial profiling is a form of profiling which comes with a steep price attached. Profiling for race will, in realistic scenarios where race correlates with crime, be considerably more costly than most other profiles, because race is invested with such importance as a marker of identity. Alienation, stigmatization and inequality can all be expected to accompany the use of racial profiling. Given this background, I have sketched three challenges, which I think combine to at once make the case for racial profiling difficult to argue, and together establish a prima facie case that racial profiling will not, in many or perhaps most circumstances, achieve minimal justification. As always when weighing consequences the balance ultimately depends on empirical facts, facts which in the case of racial profiling are extremely difficult to ascertain with any degree of certainty. In lieu of such ultimate clarification we will have to resort to intuitions and rough assessments of the strengths and weaknesses of the policies we put into practice. My suggestion is only that racial profiling, it seems, does not look promising.

If we accept this conclusion and decide to abandon racial profiling until such time as its justifiability can be better decided, the next challenge remains, for what does it mean for a police officer to not profile for race, and how would we determine whether she had or not? Setting prejudice aside, once police know that racial

differences in propensity to crime exists it may be hard if not impossible to ignore them when estimating the probability that a person is an offender. This problem is compounded by the fact that such estimations are informal or intuitive, for how will an officer know whether or not her estimate has subconsciously included race as a proxy-trait? She might try to compare her estimate with a hypothetical estimate of the same person if that person had been a different race, but such thought experiments are bound to be tricky, the conclusions vague and subject to all sorts of biases themselves – the reluctance to admit to oneself that race plays a subconscious part in one's reasoning not the least. These are very human problems, not problems restricted to instances of racism, and we should expect to find them not just in situations where officers are subject to racial prejudice, but in all police-work in communities with racially diverse populations. The real challenge, I would suggest, is therefore practical or psychological, rather than moral. The question remains how to ensure that racial profiling does not take place; how, that is, to help officers practice the art of the police, and see not that which it is useless that they should see.

Blind Justice and a Jury of Your Peers – Rescuing Procedural Legal Egalitarianism from the Egalitarians**

Introduction: What's so great about blindness and your peers?

The twin concepts of blind Justitia and a jury of your peers are staples of our representation of the justice system, but they are both relatively modern phenomena. Justitia was originally capable of seeing, her maidenhood considered sufficient guarantee of her virtues. But whether it was decided that her maidenhood had become dubious or someone eventually just realized that even virginal teenage girls are not by nature pillars of enlightened justice, a blindfold was added around the late 15th century.

The notion of trial by a jury of peers seems to originate with the Magna Carta. A free man is there guaranteed justice by a group of other freemen, as opposed to that imposed directly by the king and his loyal servants, which strikes one as an obvious benefit in cases where your interests are opposed to the crown's. Originally, however, the concept of peer was taken literally, in that your jury needed to be recruited from the appropriate social class (nobility, freemen, etc.), and it is only with the much more recent dissolution of official class distinctions that the idea comes to consist in what we now mean by it: roughly that justice is promoted by the use of a jury composed of randomly selected, average citizens.

Both notions traditionally have strong positive connotations. Yet, if we abstract from our received understanding of their metaphorical sense, these are rather strange ideas for desirable traits of the court. Certainly, as metaphorical traits go it seems preferable that the court be perceptive and astute rather than blind – unless perhaps one happens to be both the accused and guilty. And given that most of us are amateurs with respect to both the law and the methods of establishing the facts of a case, why should we desire our peers to be in charge of determining such central questions? Would the court not be better served by "a jury of your vastly better qualified superiors"? The reason that there are positive connotations attached is, I believe, that both concepts embody a particular virtue of justice, a form of impartiality often labelled equality before the law.

Yet, what is so attractive about equality before the law? A trait of the court might have many qualities that speak in favour of it. It might be truth-conducive, which would certainly be of benefit in many situations such as if it is involved in the determination of guilt and we consider the accurate answer to that question to be of at least instrumental value for the court. It might be of aid to the magistrate in establishing the appropriate punishment of those found guilty, which will presumably be important no matter which theory one subscribes to as establishing the fact of what is the appropriate punishment. It might impress on those involved the authority and competence of the court, which could be taken to serve other aims. It might even simply make the work of the court more efficient, saving everybody involved time, resources and

anguish. Intuitively, however, we do not understand the value of equality before the law to be its instrumentality to such aims, nor is it obvious that it could be. But is this form of equality of treatment in some respect intrinsically attractive, and if so why?

The tension I have alluded to in the metaphors pertain, I think, to a genuine issue of how to weigh the virtues of competency and impartiality in the context of the courtroom. In the following, however, I shall argue that this is a false opposition. The only plausible understanding of the value of equality before the law is based on conceiving of it as a norm securing minimal standards of competency – a realist nod to the difficulties that afflict actual courts, rather than an independent ambition of any ideal court. Equality before the law is instrumental to the furthering of competency in practice – no more, but also no less.

In making this case, I shall begin by establishing a number of crucial distinctions, specifically pertaining to equality in and before the law, differential treatment and likenesses. On this basis I specify a principle of procedural legal egalitarianism and the conditions of procedural legal equality obtaining in individual cases. I then examine a common objection raised against the principle of equality before the law, that it is conceptually vacuous, and argue that this is not the case. I proceed to argue, however, that on the basis of the analysis of the principle provided there is a strong case for the implausibility of a substantive moral principle of procedural legal egalitarianism, because the intuitions supporting it disappear once we remove factors supporting non-comparative principles. I examine, and reject, two potential

counterarguments, and finally suggest a way to partially rescue the principle by reinterpreting it as a consequentially grounded norm.

Let me begin, therefore with establishing the framework of the discussion by clarifying certain assumptions, defining central concepts and clarifying a number of distinctions.

Equality before the law: basic distinctions

The first and most fundamental premise for the discussion is, I believe, that the type of criminal justice system we are concerned with is at least imperfectly justified. That is, although we may allow a certain level of flaws and errors, not only as inevitable random mistakes by the system but even as structural elements, the system as a whole is sufficiently close to our requirements for a legitimate institution of criminal justice that we can consider maintaining it morally permissible.[1]

---

[1] Exactly which elements serve to legitimize and delegitimize such a system, and how wide or narrow the space for such flaws is before the system as a whole becomes unjustified, are contentious issues that I shall not attempt to resolve in the present. I believe that I can avoid doing so, because while the basic premise that the system is justified is necessary for certain steps in my argument, the specific content of the system is not. I leave to the individual reader to imagine the system in a manner compatible with her notion of what such a system must look like. I also, and with greater concern, set aside the question of what my argument would look like if we were dealing with a system which did *not* meet this condition. Here, it seems to me, it becomes much more acute exactly why and how such a system fails, because different failures will influence my argument in different ways. But despite its relevance I shall not address this aspect of the discussion in the present article.

In the following I shall take it for granted therefore that we are dealing with a system of criminal justice which serves a moral purpose that legitimizes the system, i.e. both that there are moral reasons that potentially justify punishment of persons by a system roughly similar to what we know in most parts of the world today, and that the system we are concerned with is arranged in such a manner as to gain legitimacy by adequately serving this purpose. I want to remain neutral, however, about the specific character of these reasons. While personally I hold that the foundations of a legitimate criminal justice system must be consequentialist in nature, I want to pursue the discussion in the following without taking a stand on the issue. The argument is thus meant to apply equally to those who hold other views about such foundations. Let us suppose therefore both that the criminal justice system serves a legitimate purpose – whether the allocation of just deserts, the prevention of future crime or the restoration of damaged social relations – and that the system is adequately suited to the purpose.[2]

---

[2] Some might object that an imperfect system, such as the one I allow here, can never be morally justified on consequentialist grounds given the requirements of maximising. I think this overlooks the possibility of separating the issues: it can be true at one and the same time that a system is minimally justified in that its existence produces more good than its absence would, which is what I require for the discussion at hand, and that it is sub-optimal, so that if given the opportunity we should reform it. What we are doing here is comparing different option-scenarios – one concerning status quo vs. abolition, and one concerning status quo vs. improvement. The actually right institutional design will necessarily follow from the actual options available, but for the theoretical purposes of a discussion such as the one I pursue here it will be worthwhile investigating particular sets of options, irrespective of which actual options obtain.

Further, I shall assume as minimally controversial instrumentally valuable aims of a criminal justice system 1) the *accurate* establishment of the guilt or innocence of the accused, and 2) the *appropriate* meting out of sanctions. The phrasing here is deliberately open-ended, so as to be capable of accommodating both consequentialist and retributivist understandings of the underlying purpose.[3]

Before and in the law: procedural and legislative equality

Next, it is worth specifying the focus of the discussion. Even in the context of equality before the law in an at least imperfectly justified criminal justice system there are at least two significantly different loci of equality that one could examine: the legislative and the procedural, or equality *before* vs equality *in* the law. (cf. Hart, 1997 [1961], p.159-167; Sadurski, 1986, p.131; Lippert-Rasmussen, 2010, p.171-172)

---

[3] Even some understandings of restorative justice might be compatible with the phrasing, although arguably only by stretching it to the point where it risks becoming so broad as to be meaningless. I state these points only as an assumption, however. While they seem to me highly plausible, I will not pursue an argument to that effect in the present. Similarly, abolitionists should find that these arguments apply, if their view is based on the idea that all currently realisable systems of criminal punishment will fail to properly serve the moral basis for punishment, as the second premise of our discussion here is that we are in fact dealing with such a system – they may then hold that the discussion is strictly hypothetical in that it concerns a system not currently realisable; only those that hold that no moral basis for punishment exists whatsoever will find no relevance to the issues addressed, though that should come as little surprise.

Of the two, my concern here is procedural equality. This is not because I believe the issue of equality *in* the law, that is, equality with respect to the rights and duties assigned to legal subjects by the law, to be insignificant. It poses interesting and challenging problems of its own.(see e.g. Gardner, 1989; Campbell, 1991; Gardner, 1996; Doyle, 2007) But restrictions of time and space preclude dealing with both issues at once.

What part of the justice system do I mean to encompass by procedural equality then? Roughly an element of what H.L.A. Hart called adjudicative second order rules: "Besides identifying the individuals who are to adjudicate, such rules will also define the procedure to be followed." (Hart, 1997 [1961], p.97) Based on the minimally controversial purposes I suggested above, we might say at least that the court must employ a procedure for establishing the guilt or innocence of the accused, and a procedure for establishing the appropriate sanction of the convicted. These can, in anything resembling the predominant type of legal system, be divided into smaller component procedures including e.g. establishing the facts of the case by obtaining and weighing evidence, establishing the scope and content of the pertinent parts of the law, deciding whether the established facts of the case contain a violation of the established parts of the law, and if yes, establishing the legally mandated type and scope of sanctions, the facts of aggravating or mitigating conditions, and weighing these to decide on a sanction. These can undoubtedly be further divided into smaller component procedures, but I take it that the above are sufficient to give a sense of what I take a procedure to be. Note also that we can easily

imagine alternatives to the systems we know which employ different procedures. Thus "flip a coin" is a perfectly possible procedure for establishing the guilt or innocence of an accused, although not one that we would want to substitute for the more elaborate procedure(s) that are currently employed.

Procedural equality, some might object, seems still to be concerned with rules – it is merely that the rules under consideration are those binding the actions of the court *as a court*, rather than the rules supervised and sanctioned by the court. To this I would offer two responses: first, we might acknowledge this to be true, while maintaining that there are still substantive differences between the two domains of rules and that procedural equality and its rules – formal as well as informal – are worth considering separately. Second, I should stress again that the question I am investigating is whether procedural equality – understood as a rule or not – has a moral underpinning and if so what character this underpinning has.

What I am concerned with, then, is this particular form of legal equality; the question which I shall ask is whether the notion of procedural equality can be defended as valuable independent of its potential contribution to other purposes of the system. I shall argue that this is not the case.

Treating like alike

The first salient distinction that I have presented above concerns the difference between legislative equality and procedural equality.

Whereas the first is relatively easy to grasp the second is much less easy to conceptualise than is sometimes assumed. Let me attempt therefore to clarify the notion of equality at stake. The concept of procedural equality in the law is often summed up in what is sometimes referred to as an Aristotelian principle of justice that the court must "treat like cases alike".[4] (cf. Carter, 2011, p.541; Heinrichs, 2007, p.102; Westen, 1982, p.543; Feinberg, 1974, p.310; Singer, 1978, p.186) A tentative definition might hold that an agent treats like alike in the cases of A and B *iff* the agent:

---

[4] Aristotle, fortunately given the problems I will argue the principle entails, actually says nothing of the sort. Rather, in the passage most commonly taken to support the view, he defines a particular subset of justice pertinent to the distribution of goods as characterized by comparative proportionality: "What is just will also involve at least four terms, and the ratio is the same, since the persons and the shares are divided in the same ratio. As the term A, then, is to the term B, so will C be to D, and consequently, in permutation, as A is to C, so B is to D. And so whole will bear the same ratio to whole. It is this combination which the distribution brings about, and, if the terms be united in this way, brings about justly. [...] What is just in this sense, then, is what is proportionate. And what is unjust is what violates the proportion: one side becomes too large, the other too small, which is actually what happens in practice, since the one who acts unjustly gets more of what is good, while the one treated unjustly gets less. In the case of evil, the reverse is the case, since the lesser evil is counted as a good in comparison with the greater evil; the lesser evil is more worthy of choice than the greater, what is worthy of choice is a good, and what is more worthy of good is a greater good." (Aristotle, 2000, p.86-87, 1131b-1132a) I confess that I do not find it easy to grasp exactly what Aristotle wants to argue here – the text seems to me equally capable of being read as an argument for adjusting the just allocation of goods according to desert and as an argument for distributive justice, that is, the subsuming of the distribution of goods under the issue of justice in the first place – perhaps it is both. But on any reading it concerns an issue of *telic* distributive justice, and therefore, what it certainly does not amount to, explicitly or implicitly, is an argument in favour of a deontic and procedural principle of "treating like cases alike". Joel Feinberg, it is worth noting, is at places more careful in his reading. (Feinberg, 1974, p.303, 319)

1) treats subject(s) A in manner X, and

2) treats subject(s) B in manner X

There are several problems with so simple a representation of procedural equality, however. The first is that the notion of equal versus differential treatment is more complex than the above suggests. Consider the following two cases:

> ***25% added value***: The magistrate sentences any person with trait T (T-persons) to periods of incarceration 25% longer than ¬T-persons, whenever she rules on a T-person found guilty of a crime that warrants prison.
>
> ***One size fits all***: The magistrate sentences T-persons to 6 months of prison, and ¬T-persons to 6 months of prison, whenever she rules on a person found guilty of a crime that warrants prison.[5]

Initially, it might seem that the first exemplifies differential treatment, and is thus potentially a case of procedural inequality, while the second exemplifies equal treatment, and is thus

---

[5] For ease of exposition I have chosen relatively simple examples of equal/differential treatment involving equal or different *outcomes*, but strictly speaking these are only cases of procedural inequality if we assume the diverging outcomes to be the result of different procedures being applied. It is possible, of course, that identical procedures are applied with outcomes that diverge – any procedure involving chance is capable of doing so. These ought still to be considered procedurally equal. I turn to cases involving actual procedural inequality momentarily.

potentially a case of procedural equality, but in fact neither scenario contains sufficient information to be conclusive. In the first case, it is crucial whether we assume a certain form of similarity between T-persons and ¬T-persons. If we imagine that the trait T is itself one that merits a 25% more severe punishment, suppose e.g. that T = "guilty of a crime that was coolly premeditated" and that premeditation warrants this specific increase, then it seems not to be a case of differential treatment. Similarly in the second case, if T-persons happen to be, perhaps because the trait T itself is one that entails this, persons who ought to receive sentences of less than 6 months punishment, while ¬T-persons are persons who ought to receive sentences of more than 6 months imprisonment, then this is not a case of equal treatment.

To properly exemplify differential and equal treatment we need to introduce important qualifiers:

> **25% added value v1.1**: The magistrate sentences T-persons with set of morally relevant characteristics {x, y, z} to periods of incarceration 25% longer than ¬T-persons with set of morally relevant characteristics {x, y, z}, whenever she rules on a T-person found guilty of a crime that warrants prison.[6]

---

[6] Note that, as I hope will be obvious, trait T cannot itself be part of the set of morally relevant characteristics, since I am assuming that T-persons and ¬T-persons possess identical sets of morally relevant characteristics {x, y, z} and ¬T-persons, by definition, do not possess trait T.

One size fits all v1.1: The magistrate sentences T-persons with set of morally relevant characteristics {x, y, z} to 6 months of prison, and ¬T-persons with set of morally relevant characteristics {x, y, z} to 6 months of prison, whenever she rules on a person found guilty of a crime that warrants prison.

Treatment, I suggest, is a thick concept, which presupposes a context in which certain actions are defined as relevantly similar and others as relevantly different. Another way of putting it is that while procedural equality consists simply in treating like alike, likeness is defined in a less simple manner by the morally relevant properties involved. Or, in Hart's apt formulation: "There is therefore a certain complexity in the structure of the idea of justice. We may say that it consists of two parts: a uniform or constant feature, summarised in the precept 'Treat like cases alike' and a shifting or varying criterion used in determining when, for any given purpose, cases are alike or different." (Hart, 1997 [1961], p.160)

How alike must a likeness be to be a likeness alike?

The obvious question at this point is what likenesses are and are not relevant. Note first that we are obviously not concerned with all differences of treatment. No two cases of treatment can ever be exactly the same, if only because they will necessarily differ in

either time or space.[7] Any two instances of legal procedure will differ in at least some respects. This need not violate the principle of procedural equality however, as long as these differences are trivial: that the magistrate alternately held the gavel in her right and left hand, had crossed and uncrossed legs, or spoke at a rate of 150 and 160 words per minute during proceedings cannot in any plausible scenario be held to constitute a failure of procedural equality. In effect, all procedures which differ only in morally irrelevant ways can be considered normatively equivalent variations on the same procedure. In the remainder, I refer to such a group as "procedures P". This leaves us, however, with the question of how to distinguish morally relevant differences between various procedures and various groups.

The first part of the answer is that what is morally relevant with respect to procedures and what is morally relevant with respect to groups align. This is so because morally relevant characteristics dictate which actions ought to be applied with respect to the group possessing such characteristics. The characteristics which define the group simultaneously determine the distinction between the procedure(s) that ought, and the procedures that ought not to be applied, i.e. the conclusive morally relevant difference between

---

[7] We can imagine one person being simultaneously tried in absentia by two courts using exactly the same procedures, but these cases will necessarily differ in space. Or we can imagine one person being tried twice by the same court using exactly the same procedures, but these cases will necessarily occur at different times. Any realistic scenario will, of course, involve far more, and more important, differences.

procedures. (cf. Montague, 1980, p.135; Hoffman, 1993, p.167) In short: The differences that matter are the differences that *make* a difference. Note, however, that while morally relevant differences between procedures rely on the differences between the morally relevant characteristics that groups possess, we can and should use the differences in procedure to specify the morally relevant groups. That is, we determine the morally relevant differences in procedure by the morally relevant characteristics that groups possess, but we determine the morally relevant differences between groups by the procedure(s) which ought to be applied to them.

This leaves us, however, with the deeper problem of specifying what constitutes morally relevant characteristics. The urgency of an answer to this question follows straightforwardly from the fact that we have seen it to be impossible to determine likeness of treatment independently of the likeness of the groups subjected to treatment, and impossible to determine the likeness of groups independently of the moral assessment of how we ought to treat the group in question.

However, I believe that there is a way to circumvent that issue in the specific context with which we are concerned. Thus, the second part of the answer follows from the notion of an imperfectly justified court that I have adopted. The procedure applied by such a court needs to be at least relatively consistently the morally right one. This implies that the formal and informal rules that guide its procedures can be assumed to constitute the framework for distinguishing between relevant and non-relevant traits: the relevant traits just are those which the rules of the court

take into consideration, and the reason that they are taken into consideration just is that they require differential treatment. In short: we can assume for the purposes of the present argument that the rules which mandate court procedures specify any and all relevant traits.[8]

Although this constitutes only a formal answer, this will be sufficient for the purposes of this discussion. To illustrate, consider how we could further refine the two cases introduced earlier to incorporate these considerations:

> **25% added value v1.2**: The magistrate applies procedures P to T-persons with set of morally relevant characteristics $\{x, y, z\}$ and procedures P_1 to $\neg$T-persons with set of morally relevant characteristics $\{x, y, z\}$ whenever she rules on a person found guilty of a crime that warrants prison, where P is the set of procedures the magistrate ought to apply to cases involving

---

[8] This idea faces two challenges: the first is that this is unlikely to be true in practice of any court, and the second is that it fails to provide a criterion for distinguishing between those courts or rules which do meet and those that do not meet this condition. Neither challenge, however, touches my argument in the present. Both concern the equally, if not more important, question of which principles generate the rules courts ought to apply that I have set aside in the opening remarks. Given the depth of disagreement on this issue, I prefer to remain neutral, and again proceed simply on the assumption that there is an answer to that question, and that whatever it may be, it can form the basis of the argument I present here.

persons with set of morally relevant characteristics $\{x, y, z\}$ and P_1 is not.

One size fits all v1.2: The magistrate applies procedures P to T-persons with set of morally relevant characteristics $\{x, y, z\}$, and procedures P to ¬T-persons with set of morally relevant characteristics $\{x, y, z\}$, where P are the procedures the magistrate ought to apply to cases involving persons with set of morally relevant characteristics $\{x, y, z\}$.

In fact, even these cases are somewhat oversimplified, since as noted above we can imagine a situation in which two sets of morally relevant characteristics differ, but in respects that cancel out in terms of moral requirements. That is, the two groups can possess dissimilar characteristics and still hold the same position in schemes of procedural equality, as long as the ways they differ either a) are morally irrelevant, or b) while morally relevant are different in such a way that the overall moral positions of the two groups are still identical. The latter option is an expression of the fact that all groups with moral characteristics that mandate applying procedures P constitute one group (the group of groups possessing characteristics mandating the application of P) within considerations of procedural equality.

Comparative and procedural principles

We have already moved from discussing outcomes and treatment to procedures, but the differences between these are worth briefly

emphasizing. Two distinctions in particular are worth bearing in mind. The first concerns comparative and non-comparative justice. (Feinberg, 1974) Non-comparative principles are capable of specifying what the right (or wrong) action or outcome is without reference to other outcomes or actions, although as Feinberg notes they may refer to comparative properties of the objects they apply to. Thus, it may be non-comparatively right e.g. that the fastest runner in a race (a comparative property) wins the 1st-place prize. Comparative principles on the other hand specify the proper outcome or action with reference to other outcomes or actions. Extending the example above, we could imagine that it is comparatively right in a race with prizes for the three fastest runners that the prize for the fastest runner is bigger than the prize for the second-fastest, which is bigger than the prize for the third-fastest.

On the other hand, we have the distinction between telic and deontic principles. (Parfit, 2002) Telic principles concern the goodness (or badness) of outcomes, and as a result thereof the rightness or wrongness of bringing them about, whereas deontic principles concern the rightness or wrongness of actions independent of their outcomes. And since, like Parfit, we are dealing with egalitarian variants of such principles, the principles define goodness and rightness in terms of the equality or non-equality of the outcomes or actions at stake.

Specifically, a principle of procedural legal equality will hold that applying a procedure is right if the same procedure has been applied to similar cases before. And further, it must hold, I think,

that applying the procedure is more right, the greater a proportion of similar cases have had the same procedure applied to them. If it did not hold this, but held for instance that it is simply right if the procedure has been applied to similar cases, then any exception from a practice of applying a procedure would leave future courts with equally great reasons of procedural equality to follow either the mainstream or the exception. That is, we could have a situation in which a certain procedure has been applied to all similar cases except one, and where procedural equality will now count equally in favour of applying the procedure that has been applied to almost every case and applying the procedure which was applied to the lone exception, because both will be instances of applying a similar procedure to a similar case. To avoid this, a principle of procedural equality must hold that the degree of equality matters.[9]

We might sum up the principle of **procedural legal egalitarianism** as follows:

> There is at least one reason to prefer treating a
> case with the procedures P which have been
> most frequently applied to alike cases, to treating
> it with any set of procedures which has been less

---

[9] Note that we need not assume that the reason to treat a like case like the minority or minority of like cases have been treated disappears. We can assume, perhaps more plausibly, merely that they are outweighed by the reason to treat it like the majority of like cases have been treated, but that the ratio between majority and minorities influences the strength of the reason to prefer treating the present case like the majority. This allows that only in situations where no majority exists would the reasons cancel out, which seems intuitively right.

283

frequently applied to alike cases, all else being equal.[10]

Conversely, one might hold **procedural legal non-egalitarianism**:

> Applying procedures P cannot be worse (or better) than applying another set of procedures because fewer (or more) alike cases have been treated with the same procedures P, all else being equal.

These principles borrow heavily, as I am sure is obvious, from the discussions of what Larry Temkin has dubbed "The Slogan". (cf. Temkin, 1993, p.248; Temkin, 2002; Parfit, 2002, p.98-99, 110-115) Recall however, that unlike the discussions there we are dealing with procedural equality, a subspecies of deontic egalitarianism, rather than equality of distributions in states of affairs. The focus is thus not on the comparative properties of outcomes, the locus of what Parfit labels telic egalitarianism, but

---

[10] This principle is a variant of what Derek Parfit has labelled strong egalitarianism. The moderate form would hold that though there will be such a reason, it will always be outweighed by competing reasons, that is, that we never have reason *all things considered* to prefer a situation in which more cases are treated with the same procedure, to a situation in which fewer cases are treated with the same procedure. The moderate form strikes me as an extraordinarily implausible view, requiring as it does a sort of lexical inferiority for egalitarian reasons, but I shall not engage it directly since doing so would require that we venture far into the territory of meta-ethics. (cf. Parfit, 2002, p.111-114)

on the procedures that generate outcomes. (cf. Parfit, 2002, p.88-90)

When does procedural equality obtain?

When reviewing individual cases, we want to be able to say whether the principle of procedural legal egalitarianism has been met or not. And summing up the conclusions of the section of the article, I believe we can say that **weak procedural equality** obtains in case C *iff*:

> 1) the court applies procedures P to C,
> 2) C has a set of morally relevant characteristics that mandate any one set of procedures P_x and
> 3) P is the procedure most frequently applied to cases with morally relevant characteristics that mandate procedures P_x

Note that, as should be expected from a fundamentally comparative principle, this definition is consistent with the possibility that like cases are treated alike in ways that are for other reasons morally wrong, i.e. that procedural equality obtains when no cases are treated as they ought as long as the same procedures are applied to similar cases. Note also that a benefit of the definition is that it can account for our intuitions in cases where the rules change, e.g. because parliament votes into effect an amendment to a body of law, so that there are two sets of rules mandating different procedures for the same set of cases before and after the act. Intuitively, if courts follow the rules and apply

different procedures to apparently similar cases before and after the amendment, this still ought not to constitute procedural inequality. While this might seem initially difficult to account for, it follows straightforwardly from the way that likeness is defined by morally relevant characteristics and these again by the rules of the legal system. Specifically, before and after the act there will not be two cases with morally relevant characteristics that mandate any one set of procedures P_x, but two cases which despite being similar in other respects mandate any one set of procedures P_x before the amendment and any one set of procedures $\neg P_x$ after the amendment.

Some might want to demand more of the principle of equality before the law than the above. Thus, Hart argues that for what we mean by procedural equality to be adequately captured by the maxim "treat like cases alike": "…we need to add to the latter 'and treat different cases differently'."[11] (Hart, 1997 [1961], p.159; cf. also Feinberg, 1974, p.310) We could accommodate this by adding a further condition to constrain treatment of unlike cases to being unlike. Hence **strong procedural equality** obtains in case C *iff*:

> 1) the court applies procedures P to C,
> 2) C has a set of morally relevant characteristics
> that mandate any one set of procedures P_x

---

[11] I have suggested elsewhere that this involves a conceptual albeit morally irrelevant difference between what I there label positive and negative discrimination. (cf. "Stealing Bread and Sleeping Beneath Bridges")

3) P is the procedure most frequently applied to cases with morally relevant characteristics that mandate procedures P_x, and

4) the court applies procedures $\neg P$ to all cases with sets of morally relevant characteristics that mandate any one set of procedures $\neg P_x$.[12]

Note however that the additional requirements of condition 4 are not supported by the principle of procedural legal egalitarianism. It will require a further normative principle to support it. It might be possible to supply one, but I believe the principle defined above, and the attendant conditions of weak procedural equality, to be primary to the notion of equality before the law, and as such it is on them that I shall focus the discussion in the remainder of the article.

---

[12] Some might want to go even further and require that the difference between procedures applied to unlike cases map onto the moral differences between them, so that even if improper procedures are applied to all cases no comparative inequality obtains between the ways unlike cases are treated. This would imply adding something like the following condition: 5) for any case with moral characteristics requiring procedures $P_y \neq P_x$, any difference between P_x and P is equal to the difference between P_y and the procedure applied there P_1, i.e. $P_x/P = P_y/P_1$. This condition presupposes, however, that we can measure distances between procedures cardinally, which strikes me as a rather demanding requirement. Note also that if we impose such a condition we are seemingly moving close to the ideas of the Aristotelean passage quoted above (cf. footnote 4). However, where Aristotle discusses the distribution of goods among persons according to comparative desert, we are here still concerned with the distribution of procedures across cases according to comparative moral valence. The argument remains quite different.

The Argument for superfluousness

So far, I have narrowed our focus from the broad notion of equality before the law to the more specific concept of procedural equality, and suggested that this must consist of a combination of a requirement to treat comparatively similar and dissimilar cases in particular ways as well as a method of specifying what cases fall under what headings. As noted, this seems fairly straightforwardly realisable in an imperfectly justified system of law, given that the law both defines classes of cases and specifies the required procedures to apply. The problem is, as has frequently been argued, that these enabling conditions simultaneously seem to render equality of the law superfluous. In Hans Kelsen's formulation: "And now, as for the particular principle of so-called equality before the law! It means nothing else than that the judicial institutions shall make no distinction, which the applicable law does not itself make. [...] This principle has hardly anything to do with equality. It states only that the law shall be applied as it is meant to be applied. It is the principle of legitimacy or legality, which is immanent in the essence of any legal order, regardless of whether this order is just or unjust."[13] (Kelsen, 2010 [1953], p.35-36)

---

[13] The original German reads: "Und nun gar das besondere Prinzip der sogenannten Gleichheit vor dem Gesetz! Es bedeutet nichts anderes, als daß die rechtsanwendenen Organe keine Unterschiede machen sollen, die das anzuwendende Recht nicht selbst macht. [...] Mit Gleichheit hat dieses Prinzip kaum noch etwas zu tun. Es besagt nur, daß das Recht so angewendet werden

Nor is the idea unique to Kelsen. Similar arguments can be found in Hart when he claims that: "To say that that the law against murder is justly applied is to say that it is impartially applied to all those and only those who are alike in having done what the law forbids; no prejudice or interest has deflected the administrator from treating them 'equally'. [...] The connection between this aspect of justice and the very notion of proceeding by rule is obviously very close. Indeed, it might be said that to apply a law justly to different cases is simply to take seriously the assertion that what is to be applied in different cases is the same general rule, without prejudice, interest, or caprice."[14] (Hart, 1997 [1961], p.160-

---

soll, wie es seinem Sinne nach anzuwenden ist. Es ist das Prinzip der Recht- oder Gesetzmäßigkeit, das jeder Rechtsordnung ihrem Wesen nach immanent ist, gleichgültig, ob diese Ordnung gerecht oder ungerecht ist."

[14] Note that Hart verges on the point of a potential confusion when he emphasises "prejudice, interest, or caprice". This is potentially an alluring misunderstanding, because it touches on one of the reasons why we intuitively feel that differential treatment might be morally bad: we tend to assume that differential treatment must be motivated by particular, morally reprehensible mental states directed at one of the two groups, and further that the presence of such mental states affect the moral status of the legal proceedings. This represents a confusion, it seems to me, because it conflates the qualities of the mental state of the agent with the qualities of the legal procedure. Regardless of what one believes may or may not be the relation between the mental-state of an agent and the moral status of her actions in general, it seems clear that this a separate issue from the equality of the procedure of the court. Thus, if we understand prejudice, interest or caprice to refer to the mental state of the agent, then it seems false that this will inevitably lead to an unequal application of the law. We can perfectly well suppose a prejudiced, interested and capricious magistrate, who nonetheless applies the law in the same way that an impartial, unbiased and sober magistrate would. Whatever else may or may not be morally wrong with the first of the magistrates, it strikes me as absurd to say that her behaviour would constitute the form of injustice involved in the notion of failing to treat like cases alike, given that the two magistrates *act* in the same way. Undoubtedly, it could be true if we, plausibly, allow that the deliberations of the

161) And finally Alf Ross: "There occurs not seldom in the constitutions of states a provision that all citizens are *equal before the law*. Such provisions seem void of any independent, substantial meaning. They appear to be capable of meaning only either: 1) that the law with the content that it has must be enforced without favour towards any to whom it applies – which is a truism that is already contained in the concept of a law; 2) that the law must not base its rule on such distinctions or characteristics as are considered "irrelevant" or "unjust" with respect to the legal effect at stake. But such a prohibition against "unjust" laws are devoid of any precise meaning, because "injustice" – which in the context can only mean "injustice" in a material sense – as we have seen is no more than a subjective and emotionally laden expression of disagreement towards some particular arrangement."[15] (Ross, 1953, p.371-372)

---

magistrate are part of the procedure of the court, and suppose that her deliberations are corrupted by prejudice, interest or caprice, but even this need not be true. It seems at least theoretically possible that there could be a magistrate who suffered from such flaws, but who managed to set them aside so that they did not in fact affect her deliberations.

[15] The original Danish reads: "Der forekommer ikke sjældent I staters forfatning en bestemmelse om, at alle borgere er *lige for loven*. Sådanne bestemmelser synes blottet for enhver selvstændig, håndgribelig mening. De synes kun at kunne betyde et af to: 1) at loven med det indhold den har skal håndhæves uden persons anseelse over for enhver den angår – hvilket er en selvfølgelighed der ligger allerede i begrebet om en lov; 2) at loven ikke må basere sin regel på sådanne sondringer eller kendetegn der anses for "usaglige" eller "uretfærdige" i forhold til den retsvirkning der er tale om. Men et sådant forbud mod "uretfærdige" love er uden nogen præcis mening, fordi "uretfærdighed" – der i denne sammenhæng kun kan betyde "uretfærdighed" i materiel forstand – som

That Kelsen, Hart and Ross all hold this point of view is perhaps not surprising, given their shared moral scepticism, but the analytical point can be adopted without subscribing to any of the metaethical elements of their at least partially shared normative positions.[16] The problem is the following: since what constitutes equal and unequal treatment is specified by virtue of independent and prior specifications of what the legal agent ought to do, how can a principle requiring equal treatment according to these requirements add anything to the situation?

One way of arguing that it cannot is what we might call the **argument from generality**, which holds that all rules are by their nature general, which means that they require treating all those covered by the rule in the same way.[17] In Wojciech Sadurski's words: "The principle of equal treatment of equal persons is a necessary consequence of the *general* nature of any rule which calls

---

vi har set ikke er andet end et subjektivt og følelsesbetonet udtryk for uvilje mod en vis ordning."

[16] Whether Kelsen, Ross and Hart are best understood as presupposing moral non-realism or non-cognitivism, i.e. roughly whether they individually believe that moral statements are necessarily false or simply not truth-apt, is, I believe, debatable. The finer points of exegesis in legal philosophy and metaethics need not concern us, though. Suffice to say that all three take it for granted that no moral statements accurately reflect a moral reality.

[17] My analysis in this section is indebted to Alfonso Ruiz Miguel's in "Equality before the Law and Precedent". I borrow from him the distinction between arguments for the conceptual superfluousness and the ethical superfluousness of equality before the law, as well as the distinction between an argument based on generality and an argument based on legality. (cf. Miguel, 1997, p.373-383) On all of these points, however, my analysis differs substantially from his. I return to argue against his conclusion as one of the potential counter-arguments below.

for certain treatment of certain situations. The generality of a rule consists in its application to all future cases governed by that rule. The very essence of a rule is that it brings specific situations under a general scheme; hence all equal persons (equal, that is, from the point of view of that rule's criteria of classification) must be treated in the same way. Equal treatment of equal persons is therefore nothing else but the *correct* application of a general rule." (Sadurski, 1986, p.132; cf. also Hart, 1997 [1961], quoted above; Westen, 1982, p.550-551; Winston, 1974, p.10) The argument appears to run as follows:

> the *generality of law* means that legal rules are necessarily general in nature, requiring one form of treatment by the court of the group covered by the rule,
>
> the *principle of procedural equality* requires that "equal persons (equal, that is, from the point of view of that rule's criteria of classification) must be treated in the same way",
>
> any instance following the law will, due to the *generality of law*, constitute procedural equality,
>
> no instance of procedural equality is possible outside of a context of rules subject to the *generality of law*,
>
> if a legal principle necessarily holds when the law is followed (3) and cannot hold when the law does not apply (4), then the principle is identical to following the law,

QED: "equal treatment of equal persons is therefore nothing else but the *correct* application of a general rule"

Related to but subtly different from and often conflated with the former, **the argument from legality** holds that any legal system contains a basic principle of legality, which claims essentially that the rules – and only the rules – dictate the proper working of the legal system, and that the principle of legality is prior but equivalent to the principle of procedural equality. (cf. Miguel, 1997, p.375-377) In what I take to be Sadurski's phrasing of it: "Equality in the application of legal rules means nothing more than that only differences which are relevant (from the point of view of the legal rule) should be taken into account when this rule is applied or enforced. It is the legal rule (and not, say, a judge's whim) that determines which differences are relevant. Equality before the law means, therefore, correct application of the law – and nothing more." (Sadurski, 1986, p.132; cf. also Kelsen, 2010 [1953], quoted above; Westen, 1982, p.547-549; Montague, 1980, p.136; Hoffman, 1993, p.168-170) The argument appears to run as follows:

the *principle of legality* requires (roughly) that the court act in accord with what the law dictates,

the *principle of procedural equality* requires "that only differences which are relevant (from the point of view of the legal rule) should be taken into account when this rule is applied or enforced",

whether a difference between two cases is or is not relevant is determined by whether or not there is a legal rule asserting the difference's relevance,

no instance of procedural equality is possible outside of a context of rules subject to the *principle of legality*,

any instance of following the *principle of legality* will necessarily take all and only the relevant differences into account, i.e. constitute procedural equality (by 1 and 3),

if one principle necessarily holds when a second principle is observed (4) and cannot hold when the second does not apply (5), then the former is identical to the latter,

QED: "equality before the law means, therefore, correct application of the law"

What are we to think of these arguments? Personally, I happily grant the identical premises 1), 3), and 4) of both arguments, as well as premise 5) of the argument from legality. I believe, however, that there are two important problems with the remaining premises.

The first problem is that even on their own terms both arguments fail to establish the conclusion, because 5) in the argument from generality and 6) in the argument from legality is false. This is so because on either version there could be situations in which the principle of procedural equality is respected although other

premises are not met, e.g. because persons legally defined as similar are treated equally but not according to the rules, or because although no irrelevant differences are taken into account, the court ignores certain *relevant* differences. In short, entailment is not identity.

The second, and potentially more serious problem, is that the second premise of both arguments substitutes something uncomfortably close to a mere rephrasing of the first premise's principle for a substantive principle of procedural equality – small wonder if we could derive a tautology from that.[18] Even more damaging to the argument, if replace the alternative versions of the principle of procedural equality with the substantive principle I have sketched in the opening sections of this article, then we easily encounter situations in which non-comparative principles can be followed in individual cases that simultaneously violate procedural equality. Recall that according to the principle of procedural legal egalitarianism weak procedural equality obtains in case C *iff*:

> 1) the court applies procedures P to C, and
> 2) C has a set of morally relevant characteristics
> that mandate any one set of procedures P_x and P
> is the procedure most frequently applied to

---

[18] In fairness I need to emphasize that this is a problem that extends beyond the specific quotations from Sadurski, and which concerns at bottom the failure on the behalf of critics, including Kelsen, Hart and Ross, to venture a clear definition of the concept they seek to invalidate.

cases with morally relevant characteristics that mandate procedures P_x

Clearly, as above, there can be cases where procedural equality obtains although the principle of legality is not met or the legal rules mandating general treatment are not followed, e.g. because there is a rule requiring procedure P_1 for cases with characteristic X (generality and legality), but the court applies procedure P to all such cases (procedural equality). But, and this constitutes a further problem, we can also have a case where the law is followed, so that generality and legality obtains, but procedural equality does not, e.g. because the court has previously failed to apply the non-comparatively required procedure in the majority of cases, and now switches to doing so. (cf. Miguel, 1997, p.378) In such cases, the court follows the non-comparatively required procedure, but fails to apply similar procedures to similar groups of persons.

If we take a step back and review the types of principles at stake, it is clear that the principles of legality and generality concern the relations between rules and procedures. The principle of procedural equality concerns the relation between procedures in one group of cases and procedures in another; it is comparative in a different sense. The non-comparatively right procedure is used to specify which groups are alike, or more properly speaking, the morally relevant characteristics which dictate the non-comparatively right procedure to apply are used also to specify group-likeness. As Joel Feinberg puts it: "…a non-comparative principle of justice determines the criterion of relevance for the application of the otherwise formal principle of comparative

justice for certain contexts."[19] (Feinberg, 1974, p.313) But what agents ought non-comparatively to do is not for that reason part of the definition. Procedural equality, i.e. treating like alike, is an essentially comparative principle, with different conditions for obtaining than the non-comparative principles on which it is parasitic for determining likeness. It is inevitable that they are capable of coming apart.

These problems might not overly trouble adherents of the argument for superfluousness, since it will still be true that if the non-comparative principles are consistently respected, procedural equality necessarily follows. The only slightly weaker claim would then be that even if there is not identity, and thus not conceptual superfluousness, the fact that the first principle is the more demanding of the two and that consistently respecting it will necessarily meet the requirements of the principle of procedural equality is sufficient for the *moral* superfluousness of procedural equality. Accordingly, it is to this issue that we now turn.

---

[19] Feinberg uses the specific example of a justice system which metes out non-comparatively unjust punishments in a comparatively just fashion in his discussion. (cf. Feinberg, 1974, p.312-316) For a critique of Feinberg's notion of comparative justice which to some extent mirrors my arguments in the present see (Montague, 1980, particularly p. 133; Hoffman, 1993) Montague, however, also makes claims similar to those above: "But one who acts in accord with principles of non-comparative justice will deny no one his due, and will automatically meet the requirements of comparative justice. Thus there can be no conflicts between comparative and non-comparative principles relative to actions required by the latter." (Montague, 1980, p.136) These fail, I believe, if we generalize the argument I have sketched above. However, I shall not pursue an attempt to carry out this generalization in the present, restricting my discussion to the legal context.

Moral plausibility

In the preceding, we have seen that the notion of likeness is defined *prima facie* by the legal requirements of positive law, which in themselves constitute prescriptions on the actions of the court. This implies that any case of procedural inequality will have some form of non-comparative injustice in its background. Should a magistrate apply one procedure in a case requiring a certain form of treatment and a different procedure in another which has a set of legally defined relevant characteristics mandating the same form of treatment, then she will have violated the requirements of non-comparative justice prescribing her actions in at least one of the cases.

However, we have also seen that the two requirements are capable of coming apart in individual cases, so that any individual case can instantiate either procedural equality or inequality independent of its being non-comparatively just or unjust. This gives us the opportunity to investigate the charge of moral vacuousness and establish the independent normative weight of the principle of procedural equality by showing that following a particular procedure which is procedurally inegalitarian (2 and 4 below) is morally worse than following a particular procedure which is procedurally egalitarian (1 and 3 below)[20]:

---

[20] Hoffmann and Montague both take the notion of compromising non-comparative justice in favour of comparative justice to be patently absurd and dismiss it out of hand. (Montague, 1980, p.133; Hoffman, 1993, p.173-174)

	Procedurally egalitarian	Procedurally inegalitarian
Non-comparatively just	1) Morally permissible	2) Worse?[21]
Non-comparatively unjust	3) Bad[22]	4) Worst?

Of the four possible comparisons, however, there are two that I consider impractical. Comparing 1 and 4 would tell us little, since all potentially significant moral factors change simultaneously. Meanwhile, cases comparing 1 and 2 are hard to construct, seeing as how most cases of 2 will have 3 rather than 1 as their background. However, I tentatively explore this possibility in the process of rebutting a counter-argument based on precedent below. That leaves us with two options for the present. Let us investigate first 3 vs. 4; consider:

> **The worst court in the world.** The magistrate of this appalling institution randomly selects a procedure to apply to every case. However, none of the procedures selected at random are

---

Although my arguments against it will rely on intuitions similar to those that lead Hoffmann and Montague to dismiss it, I believe that given the strong intuitions many people will have *in favour of* both equality before the law and egalitarianism in general, it may be worth exploring the issue a bit further, to illustrate how intuitions ought to count against rather than for procedural egalitarianism.

[21] E.g. the court which has previously applied procedure P_1 to cases requiring procedure P, and switches to applying P. See also examples below.

[22] E.g. the court which applies procedure P_1 to all cases requiring procedure P.

the proper procedure for any of the cases before the court, and the procedure which has been most frequently used for similar cases is never part of the pool from which the random selection occurs. Hence, every single case will violate at once procedural equality and the non-comparative requirements of applying the proper procedure.

As is hopefully apparent, this scenario involves option 4, that is, a court which is both non-comparatively unjust and procedurally inegalitarian.[23] Now, the question is this: would it be morally better if *the worst court in the world* could be reformed so as to apply always the procedure which had been most frequently applied before, and thus attain procedural equality? We must assume, that is, that instead of selecting randomly, and thus generating procedural inequality, the court always applies similar, but wrong, procedures to similar cases, which is what 3 above requires. We must further assume that in no case does this mean that a procedure that is closer to the appropriate procedure is applied, so that the functioning of the court with respect to non-comparative

---

[23] Some might object at this point that the random selection procedure constitutes a 2nd-order procedure which will, by its randomness, necessarily impose procedural equality on those subject to it; all persons are treated equally by the random procedure, that is. The example seems to me easiest to grasp as I have described it above, but it ought not to make a difference if we introduce some non-random 2nd-order procedure of selecting which of the wrong procedures to apply.

requirements is not improved. A potential objection might be that adopting procedural equality would at the very least produce prospectivity, i.e. the actions of the court would become predictable, which might have beneficial further effects. To avoid this, and similar problems, we must assume finally that the change brings with it no instrumental benefits, e.g. because the change is kept secret and the public does not realize that the system has become predictable. This assumption is necessary because such instrumental benefits are immaterial to the plausibility of the value of procedural legal egalitarianism itself.[24]

If we grant all these assumptions, then intuitively the move to procedural equality does not seem to me to make a moral difference. In fact, it strikes me as obvious that there is no reason to prefer the change to status quo. Admittedly, intuitions in such cases may be contaminated by irrelevant factors. Thus, one difficulty an argument based on intuitions will encounter is the task of separating intuitions triggered by the wrongness of violating non-comparative requirements from intuitions triggered

---

[24] Can we speak of different procedures if no improvement with respect to the non-comparative requirements is made? Does this not contradict my definition that procedures that differ only in morally insignificant ways are subsumed under a group of procedures P? I do not think so – after all, we would want to say that two procedures which differed in a morally significant way, but were both equally close to being the proper procedure, say because one does too much and the other too little in just such a manner as to make them equally bad, were in fact two *different* procedures, even if they are morally speaking equally bad.

by any wrongness which accrues to violating procedural equality. Let us, therefore, also compare across, that is 2 with 3. Consider:

> **The coin-toss court:** The court employs a procedure where anyone charged with any crime is held to be guilty, and punished accordingly, if a fair coin flipped by the magistrate comes up heads and innocent if it comes up tails. One magistrate working under this system, realizing its inherent injustice, secretly sets aside the coin and endeavours to factually determine the question of innocence or guilt. Having established this to the best of her ability she then pretends to flip the coin and proclaims a result in accord with her established convictions.

Let us assume that coin-tossing is non-comparatively the wrong procedure for establishing the guilt of the accused, and, plausibly I hope, that attempting to establish it by carefully weighing the evidence is either the right procedure to follow or at the very least non-comparatively much superior.[25] Let us also assume that the

---

[25] This is not, it is worth stressing, a requirement that she always derives the right answer, or even that counterfactually it is true that she is never mistaken in cases where the coin would have gotten it right, i.e. that she only ever makes mistakes in cases where the coin would also have produced a mistaken result. It is merely a requirement that any case decided by this process has a better probability of arriving at the right answer than the coin-toss, i.e. that the procedure is superior in accurately determining guilt and innocence. Note also that we are assuming, uncontroversially I hope, that the court is justified by its

magistrate who follows the superior procedure incurs no costs in non-comparative terms, e.g. because her procedure remains a perfect secret. Clearly, she has thus acted correctly, or at least better than she otherwise would have, according to the non-comparative principles of the court. However, in so doing she has also failed to treat like cases alike, since she treats cases correctly according to non-comparative considerations, whereas the court generally applies the coin-toss procedure.[26] Here is the question then: is there any sense in which her applying an alternative procedure, which constitutes a clear-cut case of procedural inequality, can be said to be morally bad? Does the fact that using the coin-toss will constitute procedural equality give the magistrate any reason to abandon her non-comparatively superior practice and adopt the inferior coin-toss procedure applied by her colleagues?

If we keep the necessary assumptions of all else equal in mind, then it seems perfectly obvious to me that there is no reason for the magistrate to abandon the superior procedure. The case is, as is readily apparent, a variant of the "levelling down-objection" in normative theory. (Parfit, 2002, p.98-99, 110-115) The suitably

---

purpose of ascertaining the guilt or innocence of the accused, so that the coin-toss procedure is clearly inferior.

[26] A potential objection might hold that sequence makes a difference, so that we could at once accept that cases like "repentant magistrate" do not constitute procedural inequality and that its opposite (correct procedures followed by a flawed procedure) do. I am unsure of how one might support such a claim however.

adjusted levelling-down objection supports what I labelled procedural legal non-egalitarianism, that there is never any reason to prefer applying a procedure which is inferior in terms of its meeting non-comparative moral desiderata to a procedure which is superior in these terms, simply because the inferior procedure has previously been applied in more cases, i.e. that there can never be a reason to prefer one procedure to another unless there is some non-comparative respect, in the present context defined by the purposes of the legal institution within which the procedure takes place, in which the procedure is better.

Much of the work in the debate on the original levelling-down objection is done by controversial intuitions, and I confess that I am incapable of seeing how we can avoid doing the same here. The futility of procedural equality for its own sake will be intuitively apparent or not, based on the analysis I have suggested above. However, it seems to me less controversial to suggest that equality of procedure is morally vacuous, given its independence of outcomes. In lieu of arguments to the contrary, I take it therefore that we have established a strong case against the moral significance of procedural equality. Accordingly, it is to these arguments that we now turn.

Two potential counter-arguments (and why they fail)

In the above I have argued that although not conceptually superfluous the principle of procedural legal egalitarianism is intuitively implausible, and concluded that this provides a strong case against it carrying moral significance. In the following, I will

examine two counter-arguments, the first of which is based on *stare decisis*, the principle that precedent must carry weight in the court's proceedings, and the second of which is based on the possibility of comparative disadvantage resulting from procedural inequality.

The argument from precedent

One counter-argument might hold that it is obvious both that precedent plays a role in legal reasoning and that it *ought* to play a role in legal reasoning. And if we grant this, it might be said, it follows that through *stare decisis* we have *some* reason to treat new cases like we have treated similar cases in the past, i.e. procedural legal equality, QED.

Alfonso Ruiz Miguel considers a case where non-comparative considerations presents a choice between two (or more) optimal procedures, and there is no particular reason to settle matters in one way or another, but where the court has established a precedent in favour of one procedure. Supposing that the court later reverses its initial decision and opts for the alternative procedure, it clearly violates procedural equality: "Cannot a complaint be made for reasons of equality that the second judgement has not taken sufficient account of the previous judgement (i.e. of the precedent)? [...] ...the important point seems to be that everyone must acknowledge that where there are not significant reasons for changing the interpretation of the rule already made, equality before the law remains a relevant reason [...] to hold to the precedent." (Miguel, 1997, p.382)

Taken literally, I do not think that the argument is convincing, simply because assuming there are reasons of equality that a complaint could potentially be based on begs the question – the existence of these reasons is exactly what is disputed between the proponent and sceptic of procedural egalitarianism.[27] Rather, I take it that Miguel has established an intuition in favour of following the precedent, which is capable of being explained by a principle of procedural equality. Accordingly, the argument can be countered either by defusing the intuition, or by providing a better explanation of it. I think we can do both.

As for the first, I believe the intuitive appeal Miguel has in mind trades on two factors, which are ultimately unrelated to procedural equality. First, we are inclined to suppose, indeed the example Miguel presents explicitly states, that some party will be comparatively disadvantaged by disregarding the precedent. I discuss the possibility of basing an argument for procedural legal egalitarianism on a complaint of comparative disadvantage below, and thus will not pursue it here. But to properly assess the role of precedent we must disregard it, of course, and assume the contrary: that following or disregarding precedent neither comparatively advantages nor disadvantages anyone.

---

[27] Montague, on the other hand, claims that cases like the above cannot generate a duty between comparative and non-comparative justice, because the differential treatment pertains to actions that all occur after the requirements of non-comparative justice have been met. This seems to me an implausible way of individuating the actions involved. (Montague, 1980, p.134)

Second, the example is likely to be polluted with intuitions concerning non-comparative costs of disregarding precedent. Doing so could weaken prospectivity, decrease public trust in the courts, create intra-institutional conflict, etc. All of these are legitimate concerns, and will be instrumentally important in deciding how to treat precedent. But they do not constitute a basis for arguing the intrinsic importance of procedural egalitarianism.[28] So, we must further imagine that disregarding precedent has no such costs.

Consider finally that many legal systems recognize *stare decisis*, and so will hold that precedents constitute legal sources comparable, if subordinate to, statute. In such systems, the court has a partially legislative role in addition to its adjudicative function, and in passing verdict the magistrate establishes a new legal source in addition to applying existing sources. Even legal systems which do not explicitly recognize *stare decisis* may contain informal recognition of precedents as legal sources. But in this sense, precedents are no different from legislation passed by government. The only difference is the type of agent responsible for adding to the legal sources.

It is not surprising that precedents in this aspect can give reasons for applying a procedure, since we are still assuming that they

---

[28] Such concerns might also be pertinent to deciding whether to establish a legal norm, either with respect to precedent or procedural egalitarianism directly. I discuss this possibility in the concluding section of this article.

occur in an imperfectly justified legal system, where the directives of the law correlate with what the agent ought to do. But this does not establish, of course, that past procedure, independently of its function as a non-comparatively relevant legal source, gives reasons for repeating that procedure. So we must set aside also the impact of any formal or informal rule of precedent, which establishes them as legal sources.

Conceiving of an example that meets all these requirements may not be easy. But to the extent that we can, we seem to me to be in a situation similar to that of *the Worst Court in the World* discussed above, and intuitively the case for precedent seems no stronger than in it. Consider:

> **The Amnesiac Court.** In a sophisticated future court, a neurotechnological intervention is applied to all participants in the proceedings immediately after the passing of the verdict, so as to make everyone concerned forget the proceedings. Only the outcome of the verdict is retained. Thus, no one has any recollection of the procedures followed, and no one can prefer one future procedure over another on the basis of precedent. As a consequence thereof, no formal or informal rule gives precedent the status of legal source. However, a case comes before the court which allows several equally good procedures (P, P_1, etc.) to be applied at one stage in the proceedings, all of which will

make a moral difference to the same group of people, so that no comparative advantages or disadvantages will arise from applying one rather than another.[29] Unbeknownst to all involved, all prior similar cases have applied procedures P.

In this case it seems eminently clear to me that there is no reason for the court to follow precedent. Whether it applies P or one of the alternatives makes no moral difference to the situation. What this implies is that the importance of precedent is best understood as relating to non-comparative considerations. They are non-comparative because e.g. while the costs of changing procedure are based on comparative concerns, such as the expectations of those subject to the court that future cases will mirror the past, frustrating these considerations is not.

A comparative complaint

A different counter-argument could be based on viewing the issue in terms of comparative disadvantage between the persons subjected to unlike procedures, i.e. to distinguish between whether persons are treated as they ought to be individually considered and

---

[29] Avoiding comparative disadvantage here is only possible on value-pluralist account. If we assume any value-monist account of morality, then no situation can have more than one optimal procedures P that affects the same group of people, as such sets of procedures could only differ in morally insignificant ways, and thus constitute one set of procedures P. This does not help the proponent of precedent as the basis of procedural egalitarianism however, rather, on a value-monist account no case such as the one Miguel takes as his example in favour of it could even occur.

considered in comparison to the treatment of other persons. In the preceding I have discussed non-comparatively inferior procedures loosely, without specifying in what sense, and for whom, they are inferior. But we must suppose that at least one way in which a procedure can be inferior is with respect to its treatment of the persons subject to the court, e.g. the accused. Let me focus on such types of inferior procedures in the following.

It is important to emphasize that a procedure's being inferior with respect to its treatment of the accused does not imply that the accused suffers a worse outcome. It is perfectly possible that the accused suffers exactly the same outcome as she counterfactually would have suffered had she been subjected to the non-comparatively proper procedure, or even that she suffers an outcome which is either absolutely or for her personally preferable to the outcome the non-comparatively proper procedure would have generated. The point of any procedural complaint is not concerned with the outcome, and although there could be a separate type of complaint about unequal outcomes these fall outside the scope of the present discussion. Consider:

> **Random Justice (Harsh).** A magistrate sentences all T-persons convicted of a moderately serious crime by a different procedure than the non-comparatively proper procedure, which would assign them moderately severe punishment. She instead follows a procedure of rolling dice for each of their punishments according to a system which

assigns a 40% chance of an extremely severe punishment, a 40% chance of a severe punishment, a 15% chance of the moderate punishment and a 5% chance of a lenient punishment. The dice, incredibly, consistently come out in the 15% category such that T-persons in fact receive punishments comparable to what they would otherwise have received.

In scenarios like the above, it does not make sense, of course, to say that T-persons are *disadvantaged* in terms of the outcomes – they all receive the punishments they ought to receive non-comparatively speaking. But it is still sensible to say that they are disadvantaged in terms of the sentencing procedure, because they have been subjected to a procedure that assigns them a high risk of receiving more severe punishments than the (proper) procedure used in other cases would mete out.[30] The question is this: Does suffering this disadvantage constitute a moral wrong independent of the non-comparative wrongness of the court applying an inferior procedure?

---

[30] A more complicated case would involve a procedure where T-persons are given an equal chance of getting a more lenient punishment and a more severe punishment. Supposing that the benefit of receiving a more lenient punishment is equal to the cost of receiving a more severe punishment, can the procedure be said to disadvantage them simply by imposing a lottery? My intuition is that it does not, but providing an argument to this effect takes us outside the scope of this article.

In an article on sentencing discrimination, Kasper Lippert-Rasmussen has argued that we can imagine situations in which persons who are given non-comparatively proper sentences are still in some respect sentenced wrongfully. (Lippert-Rasmussen, 2010) He describes two scenarios to illustrate how this could occur: The first parallels my example above, in that two groups of subjects are both accurately convicted (i.e. they are found to be guilty of crimes which they did in fact commit) and given the sentences that they ought to receive, but one group is convicted and sentenced through a flawed procedure which only produces the correct results through chance. The second example concerns a group which is given the sentence that they ought to receive while a similar group is given a more lenient sentence than they ought to receive.

Now, I should stress that Lippert-Rasmussen does not argue in support of what I have defined as a principle of procedural legal egalitarianism. We need to see therefore whether and if so how his argument concerning the wrongs the two groups suffer could be extended to support the principle. The first case, according to Lippert-Rasmussen, involves what on my terminology would be a non-comparative procedural wrong. Thus: "Insofar as we find sentencing discrimination unjust even in this scenario, justice in punishment must consist of something more than all criminals receiving exactly the punishment that they deserve, noncomparatively speaking…" (Lippert-Rasmussen, 2010, p.177) The wrongness, he suggests, might consist in the fact that such a legal system "fails to fully reflect the value of equality" and "sends

an objectionable message" and persist because "criminals are arguably wronged even though they receive exactly the punishment they deserve and cannot complain". (Lippert-Rasmussen, 2010, p.177) The second case, according to Lippert-Rasmussen, involves a wrong because, although the group which receives the sentence that they ought to receive is by definition treated non-comparatively justly, they are comparatively wronged. This follows from the fact that while they receive the punishment which they ought to receive (as just desert, or punitive deterrent, etc.) others who ought to be treated the same way are treated better: "Of course, it ought not to be the case that anyone receives a lighter sentence than she deserves, but given that some, but not all, do, it is fairer that everyone has an equal chance of receiving a lighter sentence." (Lippert-Rasmussen, 2010, p.179)

In combination the arguments in favour of wrong-doing in the two situations do seem to me capable of being extended to support the principle of procedural legal equality. To do so, we first need to modify our case from above so as to focus on comparative procedural disadvantage. So consider:

> **Random Justice (Lenient).** A magistrate sentences all T-persons convicted of a moderately serious crime by a different procedure than the non-comparatively proper procedure, which would assign them moderately severe punishment. She instead follows a procedure of rolling dice for each of their punishments according to a system which

assigns a 40% chance of an extremely lenient punishment, a 40% chance of a lenient punishment, a 15% chance of the moderate punishment and a 5% chance of a severe punishment. The dice, incredibly, consistently come out in the 15% category such that T-persons in fact receive exactly the punishments that they would otherwise have received.

This case parallels Lippert-Rasmussen's second case, although it involves procedural rather than outcome disadvantage. Now, if procedures are appropriate concerns of justice, as the analysis of the first case concludes, and comparative disadvantage constitutes grounds for reasonable complaint, as the analysis of the second case concludes, then it seems to be the case that comparative procedural disadvantage constitutes grounds for reasonable complaint, and that those subjected to the proper procedure (¬T-persons) might say (paraphrasing Lippert-Rasmussen): "True, I am not in a position to complain about the procedure applied to me given that it is non-comparatively just, but I am in a position to complain about being subjected to this procedure when other offenders who ought to be subjected to the same procedure are subjected to a much more beneficial procedure." (cf. Lippert-Rasmussen, 2010, p.183)

Could this extension of the argument support procedural legal egalitarianism? Not quite. Because while procedural legal inegalitarianism must involve a procedure that is comparatively inferior, it need not involve comparative disadvantage for any

individual or group of persons involved. That is, it could involve applying a non-comparatively inferior procedure to cases which are generally treated with the proper procedure, but the non-comparative inferiority could be universal, so that everyone is equally disadvantaged by the application of the inferior procedure. This might be the case e.g. if the procedure is simply inefficient in a way that unnecessarily consumes the public resources of the criminal justice system.

Presumably, many advocates of equality before the law might be willing to bite this bullet and restrict the scope of procedural legal egalitarianism to situations involving comparative disadvantage between subjects of the procedures, rather than simply between the procedures. If so, the question remains how strong a foundation the argument establishes for a scope-limited principle of procedural legal egalitarianism.

There are several potential difficulties in establishing a case for this principle. The first is that Lippert-Rasmussen presents much of the argument as a response to the claim that criminals sentenced fairly, but suffering comparative disadvantage, cannot reasonably complain, and that therefore no wrong could have been committed, and argues that it is possible for them to complain about their comparative treatment irrespective of their non-comparative treatment, and that the wrongness cannot therefore be ruled out. (Lippert-Rasmussen, 2010, p.182-184) But this, of course, does not establish the positive claim that a wrong has actually been committed, only that it cannot be ruled out on one particular ground. Putting it bluntly, we would like to know

whether there is a plausible basis for the complaint, which is equivalent to saying that we want to know whether a wrong has been committed, rather than deciding if they have plausibly been wronged by figuring out if they can complain, or else in the land of the morally wronged, Eeyore should be king. And then the case for procedural equality is thrown back on the challenge of the adjusted levelling-down objection already outlined.

Secondly, I believe that we need to carefully consider what our intuitions may be responding to about the second case. It seems obvious that all agents have grounds for complaining that T-persons are subjected to a more lenient procedure simply because they *ought* to be subject to the proper procedure (e.g. because this procedure will assess and assign just deserts). Obviously, ¬T-persons, who were subjected to the proper procedure, are moral agents, and I see no reason to think that they could not point out the non-comparative wrongness of not subjecting T-persons to the proper procedure, in the way that any other agent could. The interesting question is whether there is some special wrong involved towards these persons, that is whether their being in an in some respects comparable situation, puts them in position to be wronged by the mere fact of others being treated with a comparatively advantageous procedure. But I take it that separating our intuitions about the uncontroversial and the alleged wrong may be no easy feat.

Third, while it does seem, admittedly, that cast in terms of reasonable complaints, the fact that the magistrate treats T-persons according to an inappropriate and beneficial procedure makes a

difference for the persons who have been treated according to the proper procedure, does this mean that they have an *additional* issue to complain about? An alternative explanation could hold that rather than raising a new and separate issue, it puts them in a special position to voice the existing issue of applying the wrong procedure. That is, their comparative disadvantage serves as an additional argument for the wrongness of the procedure applied to T-persons, but not as an argument for any additional wrongs being committed. On this interpretation, the comparative disadvantage strengthens the case against applying the improper procedure to T-persons, rather than adding a claim about the comparative wrongness that could be met by applying the improper procedure to ¬T-persons, which is what would be required to support procedural legal egalitarianism. Again, it seems to me that deciding whether one or the other of these were the case would be a difficult task.

While these three difficulties hardly constitute knockdown arguments, I think they do spell out substantial challenges that an attempt to base procedural legal egalitarianism on the issue of comparatively disadvantaged groups must meet in order to provide a persuasive argument. In lieu of such, I take it we are justified in holding that procedural legal egalitarianism lacks solid grounding.

Rescuing procedural equality from the egalitarians: non-egalitarian grounds of a norm of procedural legal equality

At this point it might be tempting to echo Bernard Williams' scepticism about the moral importance of equality in general and

say of procedural legal equality simply that: "…when the statement of equality ceases to claim more than is warranted, it rather rapidly reaches the point where it claims less than is interesting." (Williams, 2006, p.231) Having reviewed, and rejected as unconvincing, the various arguments for the moral significance of procedural equality above, we are left with a choice: do we discard the principle as mere misunderstanding, or is there something yet to be said in its favour? I will argue the latter, but at the cost of shifting from intrinsic to instrumental moral significance.[31] Specifically, I will suggest that we may have good consequentialist grounds for supporting a norm of procedural equality.[32]

Norms, I take it, are socially enforced principles of practical reasoning, i.e. the rules constitutive of the decision-making procedures of an agent, and thus action-guiding but not reason-generating. Rather, norms can aid us in acting in accord with – or

---

[31] Kenneth Winston argues in a parallel way that treating like cases alike may be beneficial in particular contexts. (cf. Winston, 1974, p.36-39) The background of his analysis is substantially different from mine, however.

[32] Note that the conception of consequentialist grounds for supporting legal egalitarianism here is minimal. It involves the positive consequentialist claim that consequences matter to the moral status of an action – in this case the adoption and/or maintenance of rules of legal procedure – or putting it slightly different that agents have reasons of beneficence, not the more controversial negative claim that nothing but consequences matter. The overall argument is consequentialist in the broader sense that I have attempted to show that it is difficult to provide a persuasive explanation of an alternative principle that could influence the moral status of the relevant actions. But even those who might disagree with this first part of the argument should principally be capable of recognizing that there are beneficial consequences of adopting legal egalitarianism which speak in favour of doing so, even if *pace* my argument they believe that other moral issues have bearing on the matter.

even respond to — reasons, just as we can therefore have reasons to adopt and internalize certain norms. Note further that I take it for granted that given our psychological and cognitive constitution norms are necessarily robust. While at least some norms will presumably be subject to at least moderate revision through introspection, reflection and conscious effort, as well as due to external pressures, it is in practice impossible for an agent to review and revise her set of norms prior to any individual decision. When we assess the moral status of norms, we are by necessity assessing their quality as relatively enduring, and thus relatively general, principles of decision-making, not their quality with respect to any individual set of circumstances for a decision.

Now let us consider what might be instrumental reasons for adopting a norm of procedural legal egalitarianism. The discussion here is by necessity speculative and tentative. But although verification must remain ultimately an empirical matter, I believe a credible case can be made for assuming the beneficial consequences of a norm of procedural legal egalitarianism.

Many of the points here mirror lessons drawn from the debate on the respective advantages of act- and rule-consequentialism. No matter what one thinks of the intrinsic moral importance of rules, many act-consequentialists will acknowledge that it may be preferable if agents adopt rule-based norms as decision-making principles, rather than attempt act-consequentialist calculi.

There are several reasons for this, which can be illustrated by the comparison of "The Golden Rule" considered as a norm, vs.

"maximising good" as a norm. The Golden Rule, of course, requires that "one should do to others, as one would wish them to do to oneself". First, it is obvious that successful deliberations, i.e. those which arrive at the answer that an ideal agent would, will produce identical or very similar prescriptions in many situations, but not all. Both norms will prohibit lying, stealing, murder and adultery under normal circumstances. Assume for the sake of argument, however, both that the answers provided by the two norms differ in some situations and that act-consequentialism is true, so that all situations in which the prescriptions of the Golden Rule deviate from what act-consequentialism dictates count as a cost of applying the norm.[33] Sometimes, we are assuming, even successfully applying the Golden Rule provides the wrong answer. What are the benefits to balance (and out-weigh) this cost?

---

[33] Some might want to hold that on the best understanding of it, the Golden Rule is simply a form of impartiality, and as such cannot deviate from consequentialism. Consider a case where we might initially think the two come apart, such as killing one person in order to save two others. Here, it seems, the Golden Rule requires us to put ourselves in the place of the person whom the agent is considering killing, and will thus, unlike act-consequentialism, prohibit the killing. However, I think an argument can be made that the most plausible version of the Golden Rule will require us to put ourselves in the place of all of the affected, including the two persons who will die if the agent does *not* kill. Admittedly, the Golden Rule does not contain a principle for weighing incompatible actions where you would want one thing to be done unto you if you were some of those 'others' affected and another thing to be done unto you if you were other 'others' affected, but if all interests are assumed to count equally the Golden Rule becomes at the very least isometric with impartiality. In the present, I nonetheless assume that the Golden Rule and act-consequentialism can sometimes deviate.

First, we might suggest efficiency. Applying the Golden Rule in deliberations will, plausibly, be faster and less taxing for the agent, because the considerations it requires the agent to review are relatively few and relatively simple, compared to act-consequentialism. Act-consequentialism requires answers to the questions of what all the foreseeable consequences will be, and how much various consequences ought to count for, involving relatively uncertain intuitions about the weight of outcomes and fairly complicated probabilistic calculations to produce expected utility estimates. The Golden Rule requires merely that the agent empathizes with those affected by her actions in order to consider hypothetical scenarios in which she has traded places with them, and in which she answers the question: "Do I want her to do to me what she is doing?" In short, using the norm of the Golden Rule promises to save time and energy, which on top of the immediate saving may have derivatory benefits. Call this *the benefit of efficiency*.

Second, we might want to consider the accuracy that the norm allows the average agent to achieve. Obviously, even well-meaning agents who set aside the necessary time and energy to deliberate will not always arrive at the answer the norm would prescribe to an ideal agent, that is, they will conclude that they ought to do X, when in fact the norm(s) they base the decision upon would recommend that they do Y.[34] In fact, some norms will require

---

[34] I take it that there will be a fact of the matter in any situation as to what the agent ought to do according to some norm (as well as of course, more

321

sufficiently complex deliberations that the average agent is quite likely to fail, act-consequentialism among them. What matters, though, is not merely the percentage of errors but also the degree of error. Clearly, it is quite possible that a norm could typically produce a high percentage of errors, but errors which as mistaken prescriptions are morally almost as good as the one an ideal agent would derive. And it seems uncontroversial that such a norm could be instrumentally better than a norm which typically leads to a lower rate of errors, but with mistaken prescriptions that are vastly inferior to the one an ideal agent would derive.[35] Both the number and the degree of errors that applying a norm will typically involve thus matter to its instrumental value. Call this *the benefit of accuracy*.[36]

Third, an important feature of a norm is the extent to which it is capable of motivating the agent to take the appropriate action. I

---

fundamentally, what the agent ought to do simpliciter; as mentioned above, this imposes costs on those norms that prescribe the wrong actions even under successful deliberations), and furthermore that the ideal agent is defined as an agent who will, given a minimum of necessary conditions such as time, energy and freedom from distractions and distorting influences, deliberate so as to arrive at this answer.

[35] The observant will note that this point applies to the issue of the costs of norms deviating from the correct prescriptions mentioned above as well. And that the instrumental benefits of accuracy only apply, obviously, to norms which have sufficiently low costs. E.g. a perverse norm which correctly resulted in morally grotesque results might benefit if it was extremely complicated to apply, and therefore resulted in a high number or large degree of errors. I assume in the present that we restrict the discussion to norms with a sufficiently low cost that they avoid this problem.

[36] Strictly speaking, I am collapsing accuracy and precision into one benefit, but I take it that it is apparent from the discussion that precision is included.

take it that it is perfectly possible for an agent to successfully internalize a norm, to apply it in a choice-scenario and conclude what action it requires and yet to fail to act upon it because competing influences sway her decision. The success of various norms at motivating the agent is likely to be highly individual as well as context-sensitive, but I believe that some level of generalisation is possible here too. Thus, the fact that the golden rule has been espoused by notable religious figures and texts may make adhering to it more compelling for those with suitable religious persuasions, or even, lacking these, those brought up in a cultural environment of practices, rituals, myths and values based on a religion in which the Golden Rule features. This despite the fact that presumably whether or not an authority has advocated a moral principle ultimately provides no reason for or against acting on the principle or believing it to be valid.[37] Similarly, a norm which conflicts strongly and frequently with other influences on the agent risks being undermined. As a classic example, the demandingness of act-consequentialism conflicts strongly and frequently with the desires of almost any conceivable person, and even those agents that successfully internalize an act-consequentialist norm may find that the inevitable failure to act in accord with it most of the time robs it of motivating force. Call the

---

[37] Peter Singer claims that the Golden Rule, or presumably principles logically identical to it, can be found not only in Christianity, but also in Buddhism, Confucianism, Hinduism, Islam, Jainism and Judaism. I see no reason to doubt this claim. (Singer, 2009, p.16)

ability to motivate the agent to act in accordance with the norm *the benefit of motivation.*

Relating these to the specific context of a norm of procedural legal egalitarianism, I think it is possible that such a norm may enjoy benefits of all three types, but particularly plausible that it will enjoy benefits of the third kind.

First, as a conservative norm, which essentially advocates following established traditions, the norm of procedural legal egalitarianism may promote less deliberation of which procedure to apply. This can sound like a flaw, rather than a benefit, but bear in mind that any process of deliberation on which procedure to apply must both weigh competing considerations and reach a decision; no deliberation can usefully go on forever. By offering a simple criterion that will be easily applicable and aid in reaching resolution, the norm may promote efficiency.

Second, following past procedure will, in an imperfectly justified system, be conducive to applying the right procedure for reasons illustrated in the Condorcet Jury-theorem. Any individual agent has limited time, resources and cognitive power, but by relying on the accumulated considerations of past deliberators the agent increases her chances of deriving the right answer herself. The norm of procedural legal egalitarianism functions, on this interpretation, as a form of precautionary principle limiting individual blunders by deference to the superior capacities of the collective of past thinkers.

Third, and perhaps most importantly, the norm of procedural equality may lead to, or at the very least aid the agent in, avoiding one of the greatest challenges for any agent striving to act morally: bias. By bias, I understand the unjustified assignation of differential weights to the qualities of a person assessed by the agent, due to the personal likes and dislikes of the agent. One example would be the natural tendency for most agents to attribute greater moral worth to the persons they like, and lesser to those that they dislike.[38] But it could extend also e.g. to considerations of the trustworthiness of witnesses, estimations of how likely it is that an accused or a victim would or would not have behaved in a particular way, etc. How will the norm of procedural equality help avoid such biases? At best it seems that all it can do is recommend applying the proper procedure, which presumably is exactly what biases risk leading the agent to deviate from. How will having a separate norm requiring the same procedure be of any use?

Consider the difference between the three following requirements we might hold a magistrate to be non-comparatively subject to:

1) Apply procedures P
2) Apply only P

---

[38] The extreme here is not merely the situation where the agent unjustifiably holds a person to have no moral worth, but, at least on any normative theory which takes welfare into account, the situation where she attaches positive weight to the negative welfare of the person, and negative weight to the positive welfare, i.e. where she considers harm to the person intrinsically valuable.

3) Do not apply any procedures ¬P

Since applying P is incompatible with applying any other procedure, that is "M applies P → ¬(M applies ¬P)" it seems to follow that 2) and 3) add nothing to 1).[39] However, while morally true, this need not be the psychological case. It may be, and indeed I think it often will be, very different for an agent to respond to 1), 2) and 3), particularly in cases where biases, prejudices, preferences or interests predispose the agent to applying ¬P. The psychological force of the statements can be different, even if the moral claim of one is implied by the others. By repeating the message, the norm of procedural legal equality strengthens the influence of the requirement to apply P, and by offering a simple test of failure – did the agent apply ¬P to similar cases – it reinforces the prohibition against deviating.

Concluding remarks

In the course of the preceding, I have spelled out the notion of equality before the law in a principle of procedural legal egalitarianism, and shown that while conceptually distinct it is not a plausible moral principle. I examined two counter-arguments and illustrated why neither is convincing, but suggested that there may be a way of salvaging the principle if we understand it as a

---

[39] Recall that a set of procedures P contains all variations with trivial differences in treatment, and that a trivial difference is defined as one which is compatible with the non-comparative requirements.

consequentially grounded norm, rather than as a principle that carries independent moral weight.

The conclusion that procedural equality is morally insignificant should perhaps not come as any surprise. It fits with contentious but well-supported conclusions in the broader field of ethics that a change in the world which is not good *for* anyone cannot be good at all. But the argument I have presented here may have implications for this wider debate too. At least in the context of the legal system, I think it is possible that the intuitive support some people feel for egalitarian principles is properly attributable to the benefits of the norm. By offering an alternative explanation of the intuitions allegedly supporting one particular egalitarian principle, the argument further weakens the broader case for egalitarian moral principles.

We should keep in mind in the end, however, that a shift from intrinsic moral value to instrumentally valuable norm need not diminish the principle's practical importance. Although a court of angels wielding perfect justice would scoff at anything less than the ideal procedure, in the real world blind justice and a jury of our peers may be not merely all we can hope to get, but aspirations for which we should strive.

Jakob Holtermann, Nir Eyal, Xavier Landes, Kasper Mosekjær, Jesper Ryberg, James Shafe and Thomas Søbirk Petersen.

References

Adler, J. E. (1993). "Crime Rates by Race and Causal Relevance: A Reply to Levin." Journal of Social Philosophy **24**: 176-184.

Alexander, L. (1992). "What Makes Wrongful Discrimination Wrong? Biases, Preferences, Stereotypes and Proxies." University of Pennsylvania Law Review **141**: 149-219.

Altman, A. (2011). "Discrimination." Stanford Encyclopedia of Philosophy.

Appiah, K. A. (1994). Race, Culture, Identity: Misunderstood Connections, Tanner Lecture.

Appiah, K. A. (2006). Cosmopolitanism. London, W.W. Norton & Company.

Applbaum, A. I. (1996). Racial Generalization, Police Discretion, and Bayesian Contractualism. Handled with Discretion. J. Kleinig. Lanham, Rowman & Littlefield Publishers: 145-157.

Aristotle (2000). Nicomachean Ethics. Cambridge, Cambridge University Press.

Arneson, R. J. (1999). What, If Anything, Renders All Humans Morally Equal? Peter Singer and His Critics. D. Jamieson. Oxford, Blackwell.

Arneson, R. J. (2006). "What is Wrongful Discrimination?" San Diego Law Review **43**: 775-808.

Ayer, A. J. (2002 [1936]). Language, Truth and Logic, Dover Publications Inc.

Bagenstos, S. R. (2003). "Rational Discrimination, Accomodation, and the Politics of (Disability) Civil Rights." Virginia Law Review **89**: 825-923.

Benhabib, S. (2002). The Claims of Culture. New Jersey, Princeton University Press.

Benhabib, S. (2006). The Rights of Others. Cambridge, Cambridge University Press.

Bentham, J. (1996 [1789]). An Introduction the Principles of Morals and Legislation. Oxford, Clarendon Press.

Brest, P. (1976). "Foreword: In Defense of the Antidiscrimination Principle." Harvard Law Review **90**(1): 1-57.

Brown, C. (2011). "Consequentialise This." Ethics **121**(4): 749-771.

Campaign, F. D. A. D. (2003). "What is Discrimination." Retrieved 20 April, 2009, from http://www.stop-discrimination.info/46.0.html.

Campbell, T. D. (1991). Unlawful Discrimination. Ethical Dimensions of Legal Theory. W. Sadurski. Amsterdam, Rodopi: 153-169.

Carter, I. (2011). "Respect and the Basis of Equality." Ethics **121**(3): 538-571.

Cohen, S. (1994). "Arguing About Prejudice and Discrimination." The Journal of Value Inquiry **28**: 391-400.

Connolly, M. (2006). Discrimination Law. Suffolk, Sweet & Maxwell Ltd.

Convention for the Protection of Human Rights and Fundamental Freedoms (1950). Council of Europe.

Corlett, J. A. (1993). "Racism and Affirmative Action." Journal of Social Philosophy **24**: 163-175.

Corlett, J. A. (2011). "Profiling Color." The Journal of Ethics **15**(1-2): 21-32.

Cox, C. B. (1993). "On Michael Levin's "Responses to Race Differences in Crime"." Journal of Social Philosophy **24**: 155-162.

Crisp, R. (1988). "Utilitarianism and Vegetarianism." International Journal of Applied Philosophy **4**(1): 41-49.

Crisp, R. (2006). Reasons & the Good. Oxford, Oxford University Press.

Darwall, S. L. (1977). "Two Kinds of Respect." Ethics **88**(1): 36-49.

Darwall, S. L. (2006). The Second-Person Standpoint. Cambridge, Harvard University Press.

Dillon, R. S. (1995a). Introduction. Dignity, Character, and Self-Respect. R. S. Dillon. New York, Routledge: 1-49.

Dillon, R. S. (1995b). Toward a Feminist Conception of Self-Respect. Dignity, Character, and Self-Respect. R. S. Dillon. New York, Routledge: 290-310.

Dillon, R. S. (2009). "Respect." Stanford Encyclopedia of Philosophy.

Doyle, O. (2007). "Direct Discrimination, Indirect Discrimination and Autonomy." Oxford Journal of Legal Studies **27**(3): 537-553.

Dworkin, R. (2000). Sovereign Virtue – The Theory and Practice of Equality. Cambridge, Harvard University Press.

Edmonds, D. (2006). Caste Wars - A Philosophy of Discrimination. Abingdon, Routledge.

Enoch, D. (2009). "How is Moral Disagreement a Problem for Realism?" The Journal of Ethics **13**: 15-50.

United Kingdom. Equality Act 2006 C 3

United Kingdom. Equality Act 2010 C 15

Eyal, N. (2003). Distributing Respect. Department of Politics and International Relations. Oxford, University of Oxford. **DPhil**.

Ezorsky, G. (2001). "Discrimination." Encyclopedia of Ethics: 413-416.

Feinberg, J. (1974). "Noncomparative Justice." The Philosophical Review **383**: 297-338.

Foot, P. (1992). The Problem of Abortion and the Doctrine of Double Effect. Ethics – Problems and Principles. J. M. Fischer and M. Ravizza. Fort Worth, Harcourt Brace Jovanovich College Publishers: 59-67.

France, A. (1906 [1894]). Le lys rouge. Paris, Calmann-Lévy.

Frankfurt, H. (1997). "Equality and Respect." Social Research **64**(1): 3-15.

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

Gardner, J. (1996). "Discrimination as Injustice." Oxford Journal of Legal Studies **16:3**: 353-367.

Glasgow, J. (2009). "Racism as Disrespect." Ethics **120**: 64-93.

Glover, J. (1990). Causing Death and Saving Lives. St Ives plc, Penguin Books.

Griggs vs Duke Power Co. (1971). US Supreme Court. **401 US 424**.

Gross, S. R. and D. Livingston (2002). "Racial profiling under Attack." Columbia Law Review **102**(5): 1413-1438.

Habermas, J. (2002). The Inclusion of the Other. Cambridge, Polity Press.

Habermas, J. (2004). Between Facts and Norms. Cambridge, Polity Press.

Halldenius, L. (2005). "Dissecting Discrimination." Cambridge Quarterly of Healthcare Ethics 14: 455-463.

Harcourt, B. (2004). "Rethinking Racial Profiling: A Critique of the Economics, Civil Liberties, And Constitutional Literature, and of Criminal Profiling More Generally." University of Chicago Law Review 71(4): 1275-1381.

Hare, R. M. (1993). Essays on Bioethics. Oxford, Clarendon Press.

Harris, D. A. (2002). Profiles in Injustice. New York, The New Press.

Hart, H. L. A. (1963). Law, Liberty and Morality. Stanford, Stanford University Press.

Hart, H. L. A. (1997 [1961]). The Concept of Law. Oxford, Clarendon Press.

Haslett, D. W. (2002). "Workplace Discrimination, Good Cause and Color Blindness." The Journal of Value Inquiry 36: 73-88.

Heinrichs, B. (2007). "What is Discrimination and When is it Morally Wrong?" Jahrbuch für Wissenschaft und Ethik 12: 97-114.

Hellman, D. (2008). When Is Discrimination Wrong? Cambridge, Harvard University Press.

Hoffman, J. (1993). "A New Theory of Comparative and Non-comparative Justice." Philosophical Studies 70: 165-183.

Holst, N. (2009). Politiet finder ingen våben ved visitationer. Politiken. København, JP/Politikens Hus.

Hooker, B. (2009). Ideal Code, Real World: A Rule-Consequentialist Theory of Morality. Oxford, Oxford University Press.

Howard-Snyder, F. (2007). "Doing vs. Allowing Harm." Stanford Encyclopedia of Philosophy.

Hume, D. (2007 [1751]). An Enquiry concerning the Principles of Morals. Oxford, Oxford University Press.

Hume, D. (2008 [1739]). A Treatise of Human Nature. Oxford, Oxford University Press.

Implementing the Principle of Equal Treatment between Persons Irrespective of Racial or Ethnic Origin (2000). 2000/43/EC. The Council of the European Union.

Kagan, S. (1988). Normative Ethics. Boulder, Westview Press.

Kagan, S. (2002). The Limits of Morality. Oxford, Oxford University Press.

Kahneman, D., J. L. Knetsch, et al. (2009). Anomalies: The Endowment Effect, Loss Aversion, and Status-Quo Bias. Choices, Values, and Frames. D. Kahneman and A. Tversky. New York, Cambridge University Press: 159-170.

Kahneman, D. and A. Tversky (1974). "Judgment under Uncertainty: Heuristics and Biases." Science 185: 1124-1131.

Kahneman, D. and A. Tversky (1979). "Prospect Theory: An Analysis of Decision Under Risk." Econometrica 47(2): 263-291.

Kahneman, D. and A. Tversky (2009a). Choices, Values, and Frames. Choices, Values, and Frames. D. Kahneman and A. Tversky. New York, Cambridge University Press: 1-16.

Kahneman, D. and A. Tversky (2009b). Rational Choice and the Framing of Decisions. Choices, Values, and Frames. D. Kahneman and A. Tversky. New York, Cambridge University Press: 209-223.

Kamm, F. M. (1991). "The Doctrine of Double Effect: Reflections on Theoretical and Practicacl Issues." The Journal of Medicine and Philosophy 16: 571-585.

Kant, I. (1797). Über ein vermeintes Recht aus Menschenliebe zu lügen.

Kant, I. (1993 [1797]). On a Supposed Right to Lie because of Philanthropic Concerns. Grounding for the Metaphysics of Morals. J. W. Ellington. Indianapolis, Hackett.

Kant, I. (1996 [1797]). The Metaphysics of Morals. Cambridge, Cambridge University Press.

Kelman, M. (1991). "Concepts of Discrimination in General Ability Job Testing." Harvard Law Review 104: 1157-1247.

Kelsen, H. (2010 [1953]). Was ist Gerechtigkeit? Stuttgart, Reclam.

Kennedy, R. (1997). Race, Crime and the Law. New York, Vintage Books.

Kymlicka, W. (1996). Multicultural Citizenship. Oxford, Oxford University Press.

Law, I. (1999). "Rule-Consequentialism's Dilemma." Ethical Theory and Moral Practice 2: 263-275.

Lawrence III, C. R. (1987). "The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism." Stanford Law Review 39(2): 317-388.

Lever, A. (2005). "Why Racial Profiling is Hard to Justify: A Response to Risse and Zeckhauser." Philosophy & Public Affairs 33(1): 94-110.

Lever, A. (2007). "What's Wrong with Racial Profiling? Another Look at the Problem." Criminal Justice Ethics 26(1): 20-28.

Levin, M. E. (1981). "Is Racial Discrimination Special?" The Journal of Value Inquiry 15: 225-234.

Levin, M. E. (1992). "Responses to Race Differences in Crime." Journal of Social Philosophy 23: 5-29.

Levin, M. E. (2007). "Comments on Risse and Lever." Criminal Justice Ethics 26(1): 29-35.

Lindqvist, A. (2010). 18.000 visitationer gav 300 våben. Politiken. København, JP/Politiken.

Lippert-Rasmussen, K. (2006a). "The Badness of Discrimination." Ethical Theory and Moral Practice 9: 167-185.

Lippert-Rasmussen, K. (2006b). "Racial Profiling versus Community." Journal of Applied Philosophy 23(2): 191-205.

Lippert-Rasmussen, K. (2007a). Discrimination: What is it and What Makes it Morally Wrong? New Waves in Applied Ethics. J. Ryberg, C. Wolf and T. S. Petersen. Chippenham and Eastbourne, Palgrave Macmillan: 51-72.

Lippert-Rasmussen, K. (2007b). "Nothing Personal: On Statistical Discrimination." Journal of Political Philosophy 15: 385-403.

Lippert-Rasmussen, K. (2007c). "Private Discrimination: A Prioritarian Desert-Accommodating Account." <u>San Diego Law Review</u> **43**: 817-856.

Lippert-Rasmussen, K. (2008). "Discrimination and the Aim of Proportional Representation." <u>Politics, Philosophy & Economics</u> **7**: 159-182.

Lippert-Rasmussen, K. (2009). "Reaction Qualifications Revisited." <u>Social Theory and Practice</u> **35**(3): 413-439.

Lippert-Rasmussen, K. (2010). Punishment and Discrimination. <u>Punishment and Ethics</u>. J. Ryberg and J. A. Corlett. Chippenham and Eastbourne, Palgrave Macmillan: 169-188.

Lippert-Rasmussen, K. (2011). ""We are all different": Statistical Discrimination and the Right to be Treated as an Individual." <u>The Journal of Ethics</u> **15**(1): 47-59.

London Metropolitan Police Authority (2010). Stops and Searches Monitoring Mechanism May 2010. London, London Metropolitan Police Authority.

Loury, G. C. (2005). Racial Stigma: Toward a New Paradigm for Discrimination Theory. <u>Understanding Poverty</u>. A. Banerjee, R. Benabou and D. Mookherjee. Oxford, Oxford University Press: 654-665.

Lægaard, S. (Forthcoming). "Attitudinal Analyses of Toleration and Respect, and the Problem of Institutional Applicability." <u>Working Papers</u>, 2010, from http://www.respect.iusspavia.it/?workingpapers.

Mackie, J. L. (1990 [1977]). <u>Ethics - Inventing Right and Wrong</u>. London, Penguin.

Maitzen, S. (1991). "The Ethics of Statistical Discrimination." <u>Social Theory and Practice</u> **17**: 23-45.

McCrudden, C. (1982). "Institutional Discrimination." <u>Oxford Journal of Legal Studies</u> **2**(3): 303-367.

McCrudden, C. (1985). "Changing Notions of Discrimination." <u>Archiv für Rechts- und Sozialphilosophie</u> **21**: 83-91.

McGary, H. (1996). Police Discretion and Discrimination. <u>Handled with Discretion</u>. J. Kleinig. Lanham, Rowman & Littlefield Publishers: 131-144.

McMahan, J. (2009). Killing in War. Oxford, Oxford University Press.

Miguel, A. R. (1997). "Equality before the Law and Precedent." Ratio Juris **10**: 372-391.

Mill, J. S. (2007 [1871]). Utilitarianism. Oxford, Oxford University Press.

Miller, D. (1992). "Deserving Jobs." Philosophical Quarterly **42**: 161-181.

Miller, D. (2005). "Against Global Cosmopolitanism." The Journal of Ethics **9**: 55-79.

Montague, P. (1980). "Comparative and Non-comparative Justice." The Philosophical Quarterly **30**: 131-140.

Moore, G. E. (2000 [1903]). Principia Ethica. Cambridge, Cambridge University Press.

Moreau, S. (2010). "What is discrimination?" Philosophy & Public Affairs **38**(2): 143-179.

Morris, A. J. (1995). "On the Normative Foundations of Indirect Discrimination Law: Understanding Competing Models of Discrimination Law as Aristotelian Forms of Justice." Oxford Journal of Legal Studies **15**(2): 199-228.

Nagel, T. (1973). "Equal Treatment and Compensatory Discrimination." Philosophy & Public Affairs **2**(4): 348-363.

Nickel, J. W. (1972). "Discrimination and Morally Relevant Characteristics." Analysis **32**(4): 113-114.

Nickel, J. W. (1998). "Discrimination." Routledge Encyclopedia of Philosophy.

Nozick, R. (2010 [1974]). Anarchy, State, and Utopia. Cornwall, Blackwell.

Open Society Justice Initiative (2009). Ethnic Profiling in the European Union: Pervasive, Ineffective and Discriminatory. New York, Open Society Justice Initiative.

Parfit, D. (1984). Reasons and Persons. Oxford, Oxford University Press.

Parfit, D. (1997). "Reasons and Motivation." Aristotelian Society Supplementary Volume **71**(1): 99-130.

Parfit, D. (2002). Equality or Priority. The Ideal of Equality. M. Clayton and A. Williams. Basingstoke, Palgrave Macmillan: 81-125.

Parfit, D. (2011). On What Matters. Oxford, Oxford University Press.

Peterson, M. (2009). An Introduction to Decision Theory. Cambridge, Cambridge University Press.

Pettit, P. (1989). "Consequentialism and Respect for Persons." Ethics 100(1): 116-126.

Phillips, C. and B. Bowling (2002). Racism, Ethnicity, Crime and Criminal Justice. The Oxford Handbook of Criminology. M. Maguire, R. Morgan and R. Reiner. Cornwall, Oxford University Press: 579-619.

Pojman, L. P. (1993). "Race and Crime - A Response to Michael Levin and Laurence Thomas." Journal of Social Philosophy 24: 152-154.

Radcliffe Richards, J. (2000). Practical Reason and Moral Certainty - the Case of Discrimination. Reasoning Practically. E. Ullmann-Margalit. Oxford, Oxford University Press: 151-163.

Rawls, J. (1999 [1971]). A Theory of Justice. Oxford, Oxford University Press.

Risse, M. and R. Zeckhauser (2004). "Racial Profiling." Philosophy & Public Affairs 32(2): 131-170.

Ross, A. (1953). Om ret og retfærdighed. København, Nyt Nordisk Forlag.

Rutherglen, G. (1995). "Discrimination and Its Discontents." Virginia Law Review 81: 117-147.

Sadurski, W. (1986). "Equality Before the Law: A Conceptual Analysis." The Australian Law Journal 60(3): 131-137.

Sandberg, J. and N. Juth (2010). "Ethics and Intuitions: Reply to Singer." The Journal of Ethics 15(3): 209-226.

Scanlon, T. (2000). What We Owe to Each Other. Cambridge, Belknap Press.

Scanlon, T. (2008). Moral Dimensions. Cambridge, Belknap Press.

Schauer, F. (2003). Profiles, Probabilities and Stereotypes. Cambridge, The Belknap Press of Harvard University Press.

Schnall, S., J. Haidt, et al. (2008). "Disgust as Embodied Moral Judgment." Personality and Social Psychology Bulletin **34**(8): 1096-1109.

Segall, S. (Forthcoming). "What's so bad about Discrimination?" Utilitas.

Sidgwick, H. (1981 [1907]). The Methods of Ethics. Indianapolis, Hackett.

Simon, R. L. (1978). "Statistical Justifications of Discrimination." Analysis **38**(1): 37-42.

Singer, P. (1978). "Is Racial Discrimination Arbitrary." Philosophia **8**: 185-203.

Singer, P. (1995). Animal Liberation. London, Pimlico.

Singer, P. (2005). "Ethics and Intuitions." The Journal of Ethics **9**: 331-352.

Singer, P. (2009). The Life You Can Save. London, Picador.

Singer, P. (2010). One Atmosphere. Climate Ethics - Essential Readings. S. M. Gardiner, S. Caney, J. Dale and H. Shue. New York, Oxford University Press: 181-199.

Singer, P. (2011). Practical Ethics. New York, Cambridge University Press.

Sinnott-Armstrong, W. (2006a). "Moral Skepticism." Stanford Encyclopedia of Philosophy.

Sinnott-Armstrong, W. (2006b). Moral Skepticisms. Oxford, Oxford University Press.

Smart, J. J. C. (2008 [1973]). An outline of a system of utilitarian ethics. Utilitarianism - For & Against. Cambridge, Cambridge University Press: 3-74.

Statistics on Race and the Criminal Justice System 2008/09 (2010). M. o. Justice: 70.

Sunstein, C. R. (1991). "Why Markets Don't Stop Discrimination." Social Philosophy & Policy **8**(2): 22-37.

Taurek, J. M. (1977). "Should the Numbers Count?" Philosophy & Public Affairs **6**(4): 293-316.

Temkin, L. S. (1993). Inequality. Oxford, Oxford University Press.

Temkin, L. S. (2002). Equality, Priority, and the Levelling Down Objection. The Ideal of Equality. M. Clayton and A. Williams. Basingstoke, Palgrave Macmillan: 126-161.

Tersman, F. (2009). Moral Disagreement. New York, Cambridge University Press.

Thalberg, I. (1973). "Visceral Racism." Monist **56**: 43-63.

Thomas, L. (1992). "Statistical Badness." Journal of Social Philosophy **23**: 30-41.

Unger, P. (1996). Living High and Letting Die. Oxford, Oxford University Press.

Vallentyne, P. (2006). "Left-Libertarianism and Private Discrimination." San Diego Law Review **43**: 981-994.

Von Hirsch, A., A. E. Bottoms, et al. (1999). Criminal Deterrence and Sentence Severity. Cambridge, Hart Publishing.

Waldron, J. (1985). "Indirect Discrimination." Archiv für Rechts- und Sozialphilosophie **21**: 93-100.

Walker, N. (1991). Why Punish? Theories of punishment reassessed. Oxford, Oxford University Press.

Wasserman, D. (1996). Racial Generalizations and Police Discretion. Handled with Discretion. J. Kleinig. Lanham, Rowman & Littlefield Publishers: 115-130.

Wasserman, D. (1998). "Discrimination, Concept of." Encyclopedia of Applied Ethics **vol.1**: 805-814.

Wasserstrom, R. A. (1977). "Racism, Sexism, and Preferential Treatment: An Approach to the Topics." UCLA Law Review **24**: 581-622.

Wertheimer, A. (1983). "Jobs, Qualifications and Preferences." Ethics **94**: 99-112.

Westen, P. (1982). "The Empty Idea of Equality." Harvard Law Review **95**(3): 537-596.

Wilkinson, R. and K. Pickett (2010). The Spirit Level. London, Penguin.

Williams, B. (2006). The Idea of Equality. Problems of the Self. Cambridge, Cambridge University Press: 230-249.

Williams, B. (2008 [1973]). A Critique of Utilitarianism. Utilitarianism - For & Against. Cambridge, Cambridge University Press: 77-150.

Winston, K. I. (1974). "On Treating Like Cases Alike." <u>California Law Review</u> **62**(1): 1-39.

Woodruff, P. (1976). "What's Wrong with Discrimination?" <u>Analysis</u> **36**(3): 158-160.