

ASSESSING RACIAL PROFILING*

Steven N. Durlauf

In this article I consider the evaluation of racial profiling in traffic stops from a combination of welfarist and non-welfarist considerations. I argue that benefits from profiling in terms of crime reduction have not been identified and that further, the harm to those who are innocent and stopped is potentially high. I then argue that profiling creates a clear injustice to innocent African Americans. Together, these claims make the assessment of profiling an example of decision making under ambiguity. I resolve the ambiguity with a Fairness Presumption which leads me to reject profiling in traffic stops as a public policy.

...a cop pulled him over to the side of the road
Just like the time before and the time before that
In Paterson that's just the way things go.
If you're black you might as well not show up on the street
'Less you wanna draw the heat.

Bob Dylan, Hurricane

In this article, I attempt to develop a framework for the overall evaluation of racial profiling as a public policy. The objective of this discussion is an analysis of racial profiling that explicitly addresses the full range of implications of racial profiling for a society. In doing so, I will explicitly focus on ethical issues that arise when profiling occurs. This focus does not mean that I will ignore questions of the efficacy of racial profiling with respect to law enforcement goals. Rather, I wish to place such factors in a context in which law enforcement objectives represent a subset of the desiderata for public policy. I do not claim to provide a direct way of trading off these desiderata. Instead, I will argue that the strength of available evidence is relevant to this tradeoff; current arguments about profiling need to account explicitly for the absence of strong empirical evidence on many of the factors that come into play when assessing profiling.

The philosophical issues associated with racial profiling have recently been discussed in an important paper by Risse and Zeckhauser (2004). This paper is valuable not only for the substantive conclusions that it draws but also for carefully delineating many of the issues that need to be addressed in evaluating racial profiling as a public policy. I will follow their analysis in considering both welfarist and non-welfarist criteria¹ in assessing profiling. While I accept many of the specific

* I thank the National Science Foundation and University of Wisconsin for financial support and Lawrence Blume, William Brock, Ethan Cohen-Cole, Charles Manski, Susan Nelson, Giacomo Rondina, Wesley Skogan and Petra Todd for helpful suggestions. Two referees provided extremely helpful comments on a previous draft.

¹ Following Sen (1979) by welfarism I refer to the evaluation of policies on the basis of maximising a function that only depends (positively) on the utilities of the individuals in society. Sometimes, this is described as utilitarianism, but I reserve the latter for social welfare functions defined by the sum of individual utilities. By non-welfarist criteria, I refer to the assessments of profiling from the perspective of ethical criteria

claims in their work, I will conclude that the case for profiling is substantially weaker than they do. In the discussion that follows, I will attempt to make clear why my conclusions differ from Risse and Zeckhauser. One reason for differences in conclusions is that I evaluate racial profiling using somewhat different criteria than those employed by Risse and Zeckhauser. Another is that I restrict the discussion to a particular context. At the same time, I will question several of their arguments on their own terms.

In my discussion, I will make the following assumptions. First, I will focus on one form of racial profiling: highway traffic stops in which the objective of the stops is to identify drug carriers. As such, the pretexts for the stops (traffic violations etc.) have no intrinsic importance. Second, I assume that there are no errors in stops and searches in the sense that once someone is stopped, if he is innocent he is always let go whereas if he is guilty he is always arrested. Third, I ignore any issues of differential punishments by race; all criminals are assumed to have committed the same offence and receive the same punishment. Fourth, I do not question whether various drug laws are themselves just. By focusing on this specific context, it is possible to discuss the state of evidence on the factors that are relevant to evaluating a profiling policy. These assumptions allow the discussion to focus on the main issues specific to profiling as a police strategy. Further, by focusing on this specific context, it is possible to discuss the state of evidence on the factors that are relevant to evaluating a profiling policy.

Section 1 of this essay defines some basic probabilities concerning stops and searches, guilt and innocence, and race. These probabilities, it will be argued, are the empirical objects needed for evaluating various arguments in favour of and against profiling. Section 2 considers welfarist approaches to assessing profiling. Section 3 explores non-welfarist considerations with a specific focus on the question of fairness. Section 4 discusses the question of how to evaluate arguments in favour and against profiling when these arguments depend on quantities that are not identified from available data. Section 5 evaluates how the framework I describe affects the analysis of other policy questions, specifically profiling and terrorism and affirmative action. Section 6 contains a summary and some conclusions.

1. Basic Issues: Some Probabilities Relevant to Assessing Profiling

The basic issues that arise in evaluating profiling can be understood using simple probability arguments. This approach is developed in a recent paper by Dominitz

such as fairness. Risse and Zeckhauser (2004) divide their discussion into welfarist and deontological analyses. While my discussion of non-welfarist criteria is related to their deontological discussion, I avoid explicit discussion of deontology as my evaluation of nonwelfarist criteria is consequentialist. My reason for eschewing deontological arguments is that I wish to address tradeoffs between welfarist and nonwelfarist considerations, which seem to me easier to evaluate when one is arguing that government policies ought to promote some set of goods, as consequentialists do, as opposed to arguing that government policies must fulfil some notion of what is right as occurs in deontological argumentation. Kamm (2000) provides an introduction to deontological arguments and provides examples of this distinction between acting to achieve what is good and acting to do what is right. Petit (1991) provides what I think most economists would find to be a persuasive critique of deontological approaches.

(2003), who shows how to translate various notions of fairness in police behaviour into probability statements; I will employ a modified version of his framework and develop probabilistic characterisations for both welfarist and nonwelfarist issues. One advantage of this translation is that it helps to clarify how data matter in assessing profiling. In particular, these probabilities may be related to existing empirical studies to see how these studies inform policy evaluation. Throughout, B and W denote black and white, k denotes the fraction of the population that is black, I and G denote innocence and guilt, and S denotes the event of a police stop.

The first set of probabilities that is relevant in analysing profiling describes police behaviour. The police face a decision with respect to the allocation of a fixed rate of traffic stops C between black and white drivers. A stop strategy is a pair

$$\Pr(S|B) = \text{probability of stop if black}$$

and

$$\Pr(S|W) = \text{probability of stop if white}$$

which are consistent with the overall stop rate, i.e.

$$k \Pr(S|B) + (1 - k) \Pr(S|W) = C. \quad (1)$$

A random stop and search policy is simply a special case where the search probabilities are equal. Profiling strategies are those where $\Pr(S|B) > \Pr(S|W)$, since interest in the question derives from the oversampling of blacks.

A profiling strategy, in turn, affects the choices made by individuals on whether or not to commit a crime. Expressing behaviours in terms of innocence, these choices, which represent the only behavioural aspect of the analysis, are described by

$$\Pr[I|W, \Pr(S|W)] = \text{probability of innocence if white given stop strategy}$$

and

$$\Pr[I|B, \Pr(S|B)] = \text{probability of innocence if black given stop strategy.}$$

For simplicity I ignore sources of heterogeneity within races such as income that will lead to different criminal propensities; none of my conclusions would change if this assumption were relaxed since one could repeat the analysis conditioning on these additional factors. These race-specific probabilities should not be interpreted as implying a causal role for race.

The choice of a stop strategy combines with the innocence/guilt decisions of individuals to produce a set of equilibrium probabilities that form the basis for the assessment of racial profiling. One equilibrium probability of interest is the overall guilt rate associated with a stop strategy. I assume that all effects on the crime rate occur via deterrence, i.e. that the withdrawal of criminals from the population is too small to affect the crime rate. This allows us to express the equilibrium crime rate as

$$\Pr(G) = k \Pr[G|B, \Pr(S|B)] + (1 - k) \Pr[G|W, \Pr(S|W)]. \quad (2)$$

I will also be concerned with the rates at which innocent blacks and whites are stopped. Unlike the case of the guilt rate, these probabilities will need to be considered separately:

$$\Pr(S|I, B) = \text{probability of stop if innocent and black}$$

and

$$\Pr(S|I, W) = \text{probability of stop if innocent and white.}$$

As will be argued below, together these equilibrium probabilities capture the main objects that in principle are needed to assess profiling as a public policy. One way to think about the empirical profiling literature is to ask to what extent the existing findings may be used to identify these objects.

2. Welfarist Considerations

One approach to evaluating a profiling strategy is to do so exclusively in terms of its effect on individual welfare, i.e. utilities, for each member of the population. The racial profiling literature, in particular the empirical literature in economics, has generally not focused on questions of this type. Empirical discussions of racial profiling typically treat the number of arrests made in a set of stops as the criterion for assessing the programme. There is no reason why consistency with arrest maximisation is appropriate by itself for evaluating profiling.² I now consider some of the ways in which a profiling strategy affects individual welfare and consider the extent to which there are good reasons to believe a profiling strategy may be justified over a random stop strategy using welfarist arguments.

2.1. Benefits

What sorts of benefits might accrue to individuals from race-based stop policies? One obvious reason why profiling may produce individual benefits is its effect on the aggregate crime rate. The possibility that profiling schemes reduce crime rates lies at the heart of welfarist defences of profiling, as noted by Risse and Zeckhauser (2004, p. 144). The role of profiling in crime reduction has been formally modelled in Persico (2002) and Harcourt (2004); the basic analytical issues may be derived as follows. Suppose that the police set a stop strategy to minimise the aggregate crime rate. This is a standard optimisation problem: choose a set of stop rates to minimise (2) subject to (1). Assuming the relevant second-order condition holds and the first-order condition holds with equality, the profiling strategy that minimises the overall crime rate is one such that

$$\frac{\partial \Pr[G|W, \Pr(S|W)]}{\partial \Pr(S|W)} = \frac{\partial \Pr[G|B, \Pr(S|B)]}{\partial \Pr(S|B)}. \quad (3)$$

² This argument is also made in Persico (2002).

³ Notice that the population ratios do not appear in this equation. Intuitively, tradeoff in stop and search rates between races is exactly offset by the population ratio differences in the effects of the stop and search rates on overall crime.

An alternative way to assess the relationship between profiling and crime minimisation is to ask whether there is a profiling strategy that dominates a random stop strategy. Some form of profiling is efficient (again assuming relevant second-order conditions) when the changes in probability associated with search probabilities are unequal across races under a random profiling scheme, so that

$$\frac{\partial \Pr(G|W, C)}{\partial \Pr(S|W)} < \frac{\partial \Pr(G|B, C)}{\partial \Pr(S|B)}. \quad (4)$$

When this holds, at the margin, some disproportionate search of blacks relative to whites can be efficient.

As a theoretical matter, profiling may be required in order to fulfill a condition such as (3) or (4). This question reveals an important problem with existing empirical work on profiling, namely, the gap that exists between the current body of empirical evidence and deterrence. To assert that there are significant deterrence effects from profiling requires evidence on the sensitivity of individual crime decisions within each group to changes in the probability of being searched when guilty. However, this is not what is measured by studies that calculate the levels of crime rates across groups and there is no obvious reason why crime rates may be used to infer sensitivities to changes in arrest rates. In particular, the empirical profiling literature, in which Knowles *et al.* (2001) is a seminal contribution, has focused on the guilt rates across groups in environments where profiling occurs.⁴ The reason for this emphasis is that the main question motivating empirical studies of this type is whether the disproportionate rates at which black motorists are stopped reflect a taste for discrimination or whether they are consistent with a stop strategy that maximises the total number of arrests; the latter case means that differential stop and search rates may be interpreted as a form of statistical discrimination. A key implication of Knowles *et al.* (2001) is that arrest maximisation requires that

$$\Pr(G|S, W) = \Pr(G|S, B). \quad (5)$$

Their analysis shows that this equality in general holds for Maryland data when there is profiling of blacks in stops.⁵

However, one cannot move from this finding to conclude that observed profiling strategies are efficient in terms of reducing crime rates; (5) and either (3) or (4) are different conditions, a point also recognised in Persico (2002) using a slightly different formal framework. It is easy to imagine cases where (5) is inconsistent with either (3) or (4). As a simple example, suppose that searches are restricted to black and white teenagers and under a policing rule where the probability of a stop is independent of

⁴ The basic Knowles *et al.* (2001) framework for modelling taste discrimination versus statistical discrimination has been adopted quite widely in profiling studies in economics. A nice recent example is Anwar and Fang (2004) who study profiling in which forms of heterogeneity in motorist and policy behaviour are allowed. They concur with Knowles *et al.* (2001) that one can interpret observed stop patterns as consistent with the absence of taste discrimination.

⁵ While the Knowles *et al.* finding is consistent with the absence of taste-based discrimination, there are cases where it is consistent with the taste-based discrimination. One possibility is that black and white guilt rates are unaffected by stop rates and are further equal. Another possibility is that the stop rules for police condition on more than race, so that equal unconditional guilt rates mask differences conditional on these additional variables.

race, the percentage of black teenagers carrying drugs is higher than whites. This does not imply that a marginal change should be made to increase the search rate among blacks. It is possible that the white teenagers will exhibit greater sensitivity in their choice of whether to commit a crime to a change in detection probability than the black teenagers. The derivative of a race-specific guilt rate with respect to the stop probability will depend on factors such as stigma for arrest, which presumably is lower in communities in which high percentages of black males have been incarcerated, or because of lesser labour market opportunities.

The general reason why one cannot use the available evidence to assess overall efficiency of profiling as a crime reduction measure is that there is no one-to-one mapping between the presence of equal guilt probabilities in the presence of profiling and the efficient allocation of police effort. This is a classic identification problem in econometrics. Hence, a welfarist argument for racial profiling is difficult to sustain based on the empirical observation of equal guilt probabilities if the basis of the argument is that profiling is needed for crime minimisation. To be clear, the fact that (5) holds empirically is consistent with the claim that profiling minimises crime rates. But it is also consistent with behavioural models of criminal behaviour (i.e. descriptions of $\Pr[G|W]$, $\Pr[S|W]$ and $\Pr[G|B]$, $\Pr[S|B]$) which together represent a specification of how individuals behave in response to different stop and search probabilities) in which efficiency requires a violation of (5).

Deterrence is not the only welfarist argument for profiling. One can also argue that individuals benefit from successful arrest strategies because of the utility derived from retribution, i.e. from the fact that guilty parties are identified and presumably punished for their offenses. While one might argue that retribution is not a legitimate goal in designing a legal system,⁶ objections to it are irrelevant from a welfarist perspective, which takes the preferences of individuals as given and does not judge these preferences as to legitimacy. In other words, objections to retribution are deontological and therefore should not be considered if one is operating in a welfarist framework.⁷

In the profiling context, one can think of two ways to measure the level of retribution; no claim is made that these cover the full range of possible quantifications of the concept. First, one can equate retribution with the total number of arrests, if so, then the findings of Knowles *et al.* (2001) are consistent with retribution 'maximisation'. However, this sort of measurement seems odd, since it would imply that agents prefer a large number of arrests regardless of the number of criminals. This would require that individuals feel retribution is better served in a society with 100 criminals, 50 of whom are caught than a society in which there are 30 criminals, 29 of whom are caught. An alternative measure of retribution that addresses this concern is the number of criminals who are punished because they are caught due to a search. This measure is the

⁶ At an abstract level, I see no reason to regard retribution as an illegitimate goal of society. Otherwise, one would have to question the meting out of punishments to Axis war criminals, where the possibility of recidivism was presumably zero. Whether retribution is sensible in the context I am discussing is of course quite another matter. The point is that one needs a context to make objections to retribution.

⁷ The idea that one wishes to distinguish between deterrence and retribution as sources of individual welfare is germane if one is employing non-welfarist standards. The notion that our moral intuitions naturally differentiate between different sources of utility, to give another example, the utility derived from reading a novel versus the utility derived from watching a cockfight, is beautifully delineated in Sen (1979).

equilibrium percentage of guilty persons who are searched, $\Pr(S|G)$, which equals the ratio of the probability that a motorist is searched given one is guilty to the probability that one is guilty:

$$\Pr(S|G) = \frac{(1 - k) \Pr(S|W) \Pr[G|W, \Pr(S|W)] + k \Pr(S|B) \Pr[G|B, \Pr(S|B)]}{(1 - k) \Pr[G|W, \Pr(S|W)] + k \Pr[G|B, \Pr(S|B)]}. \quad (6)$$

Maximisation of (6) will not generally produce the condition (5). Hence the equalisation of guilt rates does not speak to whether profiling efficiently addresses retribution. Similarly, one cannot use (5) to infer whether profiling is needed for retribution. This illustrates once again that the arrest maximisation rule can fail to match welfarist objectives.

Leaving aside the difficulties of quantifying the benefits of retribution effects, there is an argument by which one might conclude that retribution is a second-order issue relative to deterrence. To the extent that deterrence and retributive aspects of profiling are in conflict, to say that one should trade them off requires that one would, at least in principle, accept a higher rate of crime in order to make sure that a higher level or percentage of the guilty are punished. It seems difficult to claim that most individuals would prefer this state of affairs. The one exception is that if punishment of a particular criminal diminishes the suffering caused by the crime (to the victim, family members, etc.) then it is possible that one would choose a higher crime rate because it minimised the number of unpunished crimes. I find this possibility sensible⁸ but do not see that it is germane to the racial profiling context, where drug crimes are not associated with individual victims.

2.2. Harm

2.2.1. Individual-specific harm from a stop and search

The consequences of a stop strategy for individual welfare extend beyond the effect of the strategy on the crime rate. In assessing the harm that profiling produces for individual welfare, it is important to consider the harm from a stop and search that accrues to a motorist. Under the assumption that the total number of stops is constant, the implications of profiling have to do with the shifting of stops from blacks to whites. In discussing this harm, I will focus exclusively on the harm to innocent blacks and whites. It is hard to imagine that the harm of a stop and search matter to a guilty party given a subsequent arrest and punishment and one can also imagine that the magnitude of feelings of humiliation and injustice from a search depend on whether the motorist is innocent. However, nothing in my argument is affected if one were to include harm to the guilty.

Are the private costs from a stop and search significant? Interestingly, this turns out to be a major question in efforts to evaluate profiling. Risse and Zeckhauser's (2004) 'in principle' defence of profiling is very much involved in arguing that these costs are small. While acknowledging that the difficulties of providing empirical support on the

⁸ One case where retribution might trump deterrence is the following. Suppose near the end of World War II a concentration camp commander offered to reduce the number of victims in exchange for immunity from post-war prosecution. The retribution costs could plausibly be high enough to reject the offer on welfarist grounds.

magnitude of the harm (p. 149) they argue that the costs to innocent African Americans from racial profiling are second-order:

... imagine how much better-off, say African Americans would be if we just got rid of profiling, *keeping everything else fixed*. We think the answer is 'only slightly so'. (p.8)

The main basis for this claim is that the harm of profiling in traffic stops only exists because of the background of past and contemporary racial harm that African Americans experience. In their view, the reaction to a search is highly sensitive to issues of personal identity and general perceptions of society.

While I concur the harm of profiling to individuals is intimately linked with background factors, in this case, the racial prejudice, I believe that Risse and Zeckhauser are quite wrong that this implies that the costs of profiling are by themselves small. There are two grounds for rejecting their claim.

First, the fact that the psychological harm to racial profiling only occurs in the context of experiences of racism and discrimination in no way implies the harm is marginal. Risse and Zeckhauser take a very particular stance on how African Americans are affected by stops and searches in a profiling environment; in essence they assume that blacks suffer from a range of racial incidents which leads to utility costs and that one more/less does not affect overall welfare very much. An alternative way to think about these costs is that background racism sensitises African Americans and so makes individual incidents such as search and stops much more harmful to their wellbeing than would otherwise be the case. By analogy, the pain of slapping my back is far greater when I am sunburned than when I am not. So, while the effect of the slap is almost entirely contingent on the sunburn, it is because of the sunburn that the pain is severe. The strength of opposition of African Americans to profiling at least hints that my interpretation is more likely the correct one.

In fact, it is difficult to reconcile Risse and Zeckhauser's views on the interactions of race and profiling in producing costs with their claim that the effects of profiling are, *ceteris paribus*, second order. Risse and Zeckhauser's most precise description of the nature of the utilitarian harm to profiling is that 'We believe that the harms of racism and profiling may be subadditive' (p.149). To make sense of this remark, suppose that the cost function for racial incidents for an individual is $c(P,x)$ where P is the measure of profiling incidents and x a measure of background racism. Assume that $c(0,0) = 0$; this a normalisation and does not matter. Subadditivity means that $c(P,x) \leq c(0,x) + c(P,0)$ which implies $c(P,x) - c(0,x) \leq c(P,0) - c(0,0)$. But this inequality means that the harm from profiling is decreasing in the level of background racism, which is inconsistent with their claim that profiling is only harmful because of background racism. An alternative formulation is that profiling incidents and background racism are additive, i.e. $c(P+x)$ where $P+x$ equals the total number of racial incidents. Subadditivity would in this case mean that $c(P+x) \leq c(P) + c(x)$ or, normalising so $c(0) = 0$, that $c(P+x) - c(x) \leq c(P) - c(0)$. In this case, the cost function is concave in the scalar measure. The claim that on the margin, profiling has second order effects on costs, seems more sensible for this case (although again, it is still no longer clear what it means that the harm from profiling is because of background racism, since the level of background racism reduces the marginal harm of profiling).

However, this formulation eliminates any notion that racial background and profiling are distinctive determinants of harm and is inconsistent with the rest of their discussion. The problem with the Risse and Zeckhauser formulation of interactions between profiling and racism and overall harm is that they appear to confuse the concavity of costs with respect to a single racist measure (which is a statement about a second derivative) with the interaction of profiling with background, which is a statement about a cross-partial derivative.

Second, the assumption that a stop and search is intrinsically a minor inconvenience is questionable. There is no particular reason to believe this is the case in practice. Gross and Barnes (2002, pp. 745–6) provide the following description of a traffic stop made on the pretext of allowing a search for drugs:

As the level of the police officer's interest increases, the cost to the innocent citizen escalates rapidly. It's one thing to get a speeding ticket and an annoying lecture ... it's quite another to be told to step out of the car and to be questioned ... The questions may seem intrusive and out of line, but you can hardly refuse to answer an armed cop. At some point you realize you are not just another law-abiding citizen who's being checked out ... like everyone else. You've been targeted. The trooper is not going through a routine so he can let you go ... he *wants* to find drugs on you ... Those of us who have not been through this sort of experience probably underestimate its impact. To be treated as a criminal is a basic insult to a person's self-image and his position in society. It cannot easily be shrugged off ...

The Gross and Barnes description is not one which describes aberrant or abusive behaviour by the police. What they describe is typical behaviour when a police officer has assigned some chance to a driver possessing drugs; and of course underlying this interaction is presumably some fear on the part of the officer that, if guilty, the driver is dangerous.

One can of course further well imagine these interactions becoming abusive. Risse and Zeckhauser (2004) are explicit in evaluating profiling when there is no abuse on the part of the police. But no argument is made that this is possible. In terms of evaluating profiling in traffic stops, this assumption is problematic, since any evaluation of racial profiling needs to account for what is feasible. By analogy the claim that the best government is a benevolent monarchy, since by definition it will choose policies that are best for society and implement without impediments, is of little interest since such a government could not exist. Hence, unless Risse and Zeckhauser can make an argument that non-abusive profiling is possible, I do not see that this assumption is tenable. One can just as easily argue that profiling creates negative stereotypes in the minds of the police and makes abuse more likely. Risse and Zeckhauser could of course ask why the burden of proof is on them to prove non-abusive profiling is possible, as opposed to the burden lying with me to prove it is not. The answer to who bears the burden of proof in such cases is part of my discussion in Section 4.

Beyond the question of the magnitude of the costs of profiling in isolation, there are also good reasons to question whether a conclusion that they are small necessarily translates into a statement that they should have little effect on evaluating profiling as a policy. These arguments do not question the Risse and Zeckhauser claim that profiling

costs in isolation are small, but rather suggest that the translation of their claim into a policy conclusion about profiling in traffics stops works in favour of profiling in policy evaluation.

One issue concerns the definition of a policy evaluation. Even if the marginal effect of profiling on each individual is small because of the background of racism, this does not mean that the policy is defensible in terms of its costs, unless one restricts defensibility in an unappealing way. Suppose that the harm to African Americans comes from a set of two types of experiences with others and that the harm accrues if either type occurs. Let one experience be a (non-abusive) stop and search which occurs under a profiling regime and the second any other government policy that is currently in place. If one is eliminated and the other is not, then the harm to the African American will not be reduced. However, for larger sets of potential public policy changes i.e. simultaneously eliminating profiling and the second government policy, the reduction in costs will be high. The fact that one may need to bundle elimination of racial profiling with other actions does not constitute a welfarist argument that its costs are unimportant.

Further, a conclusion that the harm is small is in principle consistent with a welfarist calculation that the harm is significant. Suppose that social welfare is defined by

$$SWF = \sum_i h(u_i), \quad (7)$$

where u_i is individual i 's utility. If $h(\cdot)$ is concave, then social welfare will reflect decreasing marginal benefits to changes in the utility of the low utility agents versus high utility agents. Now suppose that Risse and Zeckhauser are correct that blacks have low utilities due to the pervasive background of prejudice they experience. If so, then sufficient concavity of $h(\cdot)$ can produce the result that further decreases in the utility of blacks cannot be justified through increases in the utility of others, even if the changes in black utility are, when considered in isolation, 'small' compared to the increases for others. Of course, this all depends on the sizes of the benefits, which I have argued are not known.

2.2.2. *Social harm*

Beyond the direct harm of stops to innocent motorists, it is possible that there is social harm that is associated with profiling strategies that needs to be accounted for in assessing the effects of profiling on individual welfare. More generally, when one considers the effects of profiling in the broader social context, one can identify costs beyond the private ones associated with a stop. In a profound recent study of racial inequality in America, Glenn Loury has argued that persistent inequality between blacks and whites may be understood as stemming from the effects of stigma on blacks. Loury (2002, p. 9) defines stigma as

... the identity unreflectively imputed to someone by observers who, not being privy to extensive idiosyncratic information, draw conclusions about a person's deeper qualities on the basis of easily observable indicators that may lie at hand.

In my interpretation of Loury's argument, stigma generates racial inequality because members of a society face a fundamental identification problem in evaluating issues

associated with race, given the limited information from which such evaluations are made. Beliefs such as ‘blacks are less intelligent than whites partially due to genetic factors’ may be nonsense from the perspective of the body of evolutionary biology but do not lend themselves to refutation given selective and limited observations on African Americans. Further, such beliefs can be self-perpetuating. If stereotypes about racial differences in intelligence contribute to lower educational attainment by African Americans, by discouraging African Americans from educational investment, producing disidentification of the type studied by Claude Steele (1997) and others, then the beliefs can be reinforcing.⁹

In this context, racial profiling may be argued to contribute to the promotion of racial stigma. By treating race as an appropriate criterion for policing decisions, the perception that crime and race are ‘fundamentally’ linked is reinforced. The potential deleterious effects of thinking in categories have a long tradition in social psychology. The classic Robbers Cave experiment (Sherif *et al.*, 1961), in which adolescents who were divided arbitrarily into groups developed intergroup prejudices and hostility is the most famous example of such research. The behavioural consequences of stereotyping have also been well documented; one interesting controlled experiment due to Rogers and Prentice-Dunn (1981) shows how angered whites will be more aggressive towards blacks than whites for the same ‘offence’. Further, the perception of injustice in society by African Americans can be reinforced by the emergence of stigma, which will increase the costs of profiling.

Of course, there is no body of evidence that quantifies how racial profiling affects levels of racial stigma or how racial stigma affects African Americans. This of course does not mean that the costs should be assumed to be negligible. Rather, it indicates the difficulty in formulating a welfarist decision problem, an issue that I return to below.

2.3. *Profiling as Redistribution*

In considering the different arguments on benefits and harm to profiling, it is worth noting that it is widely recognised that African American communities do not support racial profiling by the police. This means (if one takes these preferences as fixed) that it is highly likely that in a welfarist calculation, there will be tradeoffs between the welfares of blacks and whites under profiling. Since the Pareto improvement arguments cannot be invoked, one is left with the question of how to evaluate such tradeoffs. What makes such an assessment difficult is that one is in essence trading off the private harm to stops as well as stigma and respect for one group against lower susceptibility to crime (which will apply to both groups). Hence, any welfarist argument for or against profiling will be sensitive to any implicit egalitarianism built into the social welfare function.

⁹ Notice that this is a somewhat different claim from standard formulation of statistical discrimination. In statistical discrimination models, beliefs about the stigmatised group are confirmed *ex post* in equilibrium. Here, the false beliefs produce outcomes that militate against their refutation, possibly due to identification problems. This is one reason why I believe Loury’s notion of racial stigma is an important advance on statistical discrimination as an explanation of racial inequality.

2.4. *Evaluating Welfarist Claims*

I conclude that there is no strong welfarist argument for profiling. Does this contradict Risse and Zeckhauser (2004), who argue that in principle a welfarist justification for profiling does exist? While Risse and Zeckhauser agree that the case for profiling in traffic stops is weak due to a lack of evidentiary support, they present their analysis as a defence of profiling in some cases. In my view, their abstract argument is, given their assumptions, clearly correct. If one posits first, a welfarist objective and second, the possibility that the costs of profiling are plausibly small relative to the benefits, measured in terms of individual utilities, then it is necessarily the case that profiling may be justified in principle. However, I would argue that for the case under primary public policy dispute, traffic stops, these assumptions are highly questionable. There is no principled way to assign probabilities to the asserted costs and benefits since the empirical literature identifies neither deterrence effects nor individual and social harm. And to the extent to which one relies on ‘fuzzy’ notions of the likelihoods of certain levels of costs and benefits, I disagree with qualitative claims made by Risse and Zeckhauser, notably on the likelihood that costs are small. This leads me to a different conclusion from Risse and Zeckhauser (2004). However, to say the case is weak does not mean that in an expected value sense, the net welfare benefits to profiling are low. What it means is that many of the factors that determine the welfare effects of profiling are not known to a policy maker. Whether this ultimately leads to a rejection of profiling will require some additional argumentation on how to assess ‘ambiguous’ environments and is taken up in Section 4.

3. Non-welfarist Considerations

By non-welfarist considerations, I refer to evaluations of profiling that are based upon standards that lie outside the specification of a social welfare function whose elements are individual utilities. Specifically, in evaluating states of social affairs and the effects of government policies on these states, I wish to allow for explicit consideration of ethical notions, in particular consideration of fairness, which may function independently of the welfarist calculus. In doing so, I follow an approach advocated by Sen (1979) that is consequentialist in that it evaluates policies according to the states of affairs they induce rather than in terms of deontological restrictions on permissible choices, yet that allows the evaluation of consequences according to criteria beyond their effect on personal utility; as Sen (1979) writes,

... considerations of liberty and rights have been viewed here as parts of the structure of outcome morality itself ... This contrasts with treating them as constraints on, or nonconsequentialist judgments of, actions ... This shift is possible because of the departure from the tradition – often implicit – of identifying *consequences* with utility consequences (and basing the description of states of affairs entirely on utility information regarding those states).

Hence, even if one sees a plausible welfarist argument in favour of profiling, one may reject profiling because of its inconsistency with these other factors.¹⁰

Non-welfarist arguments may be subjected to the criticism that they may lead to violations of the Pareto principle. Sen (1970), for example, shows how the Pareto principle may conflict with individual liberties in a way that there exist configurations of laws that restrict liberties yet are unanimously preferred to any alternative in which these liberties are preserved. Kaplow and Shavell (2001) make a broader argument that considerations such as a concern for fairness may lead to conflicts with the Pareto principle.

My view is that there is no reason why the Pareto criterion should be regarded as having some special, i.e. lexicographic, ethical standing. This is not to say that welfarist considerations should not be a primary consideration in evaluating policies, only that non-welfarist considerations may play a primary role as well. Arguments in support of the primacy of the Pareto criterion imply a prior ethical judgment that there is no feature of social good outside the utilities of society's members. There is no reason why one need reject the existence of these other features. This idea motivates much of Sen's writings. Hahn (1982) p. 188) makes the argument as follows, in the context of the value of liberty:

A social state is not fully described by me if I am only given the utilities of the agents in that state. I also need to know the liberty enjoyed by them. It follows that my ranking of social states cannot be of the form of the social welfare function whose arguments are only the utilities of individuals. If the utilitarian asks me why I should care about liberty over and above what is already recorded in the utility functions, I can answer that, for me, liberty is an intrinsic good just as for him utilities are intrinsic goods.

Hahn's point is that in evaluating a state of affairs for society, in addition to considering the utilities for individuals, he wants information on other properties of the state. While he mentions the way in which the state embodies personal liberty, one can generalise this idea to include a concern for justice or other qualities that are not summarised by the utilities of individuals.¹¹

To make this view concrete, suppose that an all-white society passed a set of laws that explicitly discriminated against blacks. Is it reasonable for someone outside the society to argue against them on the basis that they are unjust? I believe that the answer is clearly yes even though

- (1) no discriminatory consequences occur to any member of the society and
- (2) individuals may derive some utility from the existence of the laws.¹²

Similarly, one could argue that one society is to be preferred relative to another due to the level of scientific and cultural attainments even if individuals in the other society

¹⁰ Fairness is of course not the only non-welfarist standard. Risse and Zeckhauser (2004) consider, from a deontological perspective, whether profiling policies may be rejected because they are discriminatory. Their argument against labelling such policies as pejorative discrimination is, in my view, very persuasive.

¹¹ The idea that multiple conflicting social goods exist appears in many contexts; for example it is the basis of Isaiah Berlin's writings on value pluralism, e.g. Berlin (2002).

¹² This does not mean that the moral offensiveness of these is independent of the scope of their effect. By analogy, contemporary anti-Semitism in Japan is less offensive than anti-Semitism in France at least partially because of its irrelevance; my argument is simply that the discriminatory laws are unethical and that this is a basis for rejecting the laws.

are happier. These examples suggest why one might in principle decouple ethical considerations from unanimity. As noted above, following Sen (1979), one can make elaborate arguments that factors outside of individual welfare are morally relevant in evaluating societal outcomes, even if one places positive weight on individual welfare.¹³ An appealing feature of Hahn's formulation is that it recognises that nonwelfarist considerations do not have to trump individual welfarist considerations but that both are relevant in assessing policies.

In the context of profiling, the considerations that impinge upon any welfare calculation are those that derive from our notions of justice. Here, I will focus on a particular notion of fairness, one that may be linked to some of the conditional probabilities I have defined in Section 2. To do this, I argue that an appropriate fairness criterion for profiling may be derived from the notion of equality of opportunity, specifically as understood by John Roemer in a series of studies (Roemer, 1993, 1998).¹⁴ Roemer makes the following argument. Suppose that society is considering an outcome such as education, and wants to determine whether the society provides equality of opportunity with respect to that outcome. Roemer argues that to do this, one must identify the determinants of the outcome and divide them into two categories: determinants for which an individual should be held responsible and determinants for which the individual should not be held responsible. A society should act in some way to indemnify individuals against harm that accrues due to those factors that they cannot control.¹⁵

Roemer's argument may be interpreted probabilistically as implying that equality of opportunity requires that the conditional probability of an outcome should only depend on those factors for which an individual is responsible.¹⁶ This general idea has straightforward application for the profiling problem since innocence and guilt are clearly characteristics for which an individual should be held responsible whereas race is clearly one that is not. In the profiling context, I will not refer to equality of opportunity but to fairness. For an innocent individual, complete fairness implies that the conditional probability of the negative outcome of being stopped should not depend on his race,¹⁷

¹³ Without going into detail here, many of the objections to welfarism amount to (1) rejecting that idea that a policymaker cannot make judgments about the 'legitimacy' of individual preferences and (2) the view that there are intrinsic reasons that justify the pursuit of properties such as procedural justice that function outside of welfare considerations.

As Sen (1979) states:

... a tortured body, an unfed belly, a bullied person, or unequal pay for equal work, is as much a part of the description of states of affairs as the utility and disutility occurring in that state.

¹⁴ Analyses such as Roemer's represent the modern philosophical effort to provide foundations to egalitarianism. Cohen (1989) is an excellent analysis of different approaches to egalitarianism, and well summarises the egalitarian idea:

A person is *exploited* when unfair advantage is taken of him and he suffers from (bad) *brute luck* when his bad luck is not the result of a gamble or risk he could have avoided. I believe that the primary egalitarian impulse is to extinguish the influence on distribution of both exploitation and brute luck.

Cohen's notions of exploitation and brute luck are both involved in my discussion of fairness. Decomposing their respective roles is perhaps a useful subsequent exercise.

¹⁵ Roemer recognises that this distinction is not self-evident and needs to be adjudicated as part of a political process.

¹⁶ I defend this interpretation in Durlauf (1999, 2002).

¹⁷ There is no guarantee that complete fairness can hold in a given context. Measures of the extent of violation of (8), for example $|\Pr(S|I,W) - \Pr(S|I,B)|$, can, in principle, be used to assess degrees of unfairness for alternative policies.

$$\Pr(S|I, W) = \Pr(S|I, B). \quad (8)$$

What implications does the fairness requirement (8) have for racial profiling strategies? This equality may be rewritten

$$\frac{\Pr(S|W) \Pr(I|S, W)}{\Pr(I|W)} = \frac{\Pr(S|B) \Pr(I|S, B)}{\Pr(I|B)} \quad (9)$$

or

$$\frac{\Pr(S|W)}{\Pr(S|B)} = \frac{\Pr(I|S, B)}{\Pr(I|B)} / \frac{\Pr(I|S, W)}{\Pr(I|W)}. \quad (10)$$

In this expression, $\Pr(I|S, W)/\Pr(I|W)$ and $\Pr(I|S, B)/\Pr(I|B)$ represent the ratios of the conditional probabilities of innocence given race and search to the probabilities of innocence given race. Therefore, it is possible for racial profiling to co-exist with fairness in the treatment of the innocent. However, any differences in search rates that are race-based are consistent with fairness if and only if they reflect differences in efficiency of search decisions, i.e. that for the race where the stops occur more frequently, there is better screening of the innocent from others. If there is no such information, then, each of these ratios equals 1 and fairness implies that regardless of the underlying innocence probabilities, the conditional probability of stops and searches should not differ across race, i.e. equal accuracy in assessing guilt implies that stops and searches should be random with respect to race, i.e.

$$\Pr(S|W) = \Pr(S|B). \quad (11)$$

If one accepts my definition of fairness, then it is straightforward to see why racial profiling may engender a basic tradeoff between fairness and efficiency. As illustrated by a comparison of (8) with (3), the fairness criterion may be inconsistent with targeting stops and searches in order to minimise the crime rate; the criterion is also inconsistent with (5), the criterion that differentiates statistical from other types of discrimination.

Do existing racial profiling practices in fact involve a violation of fairness? It appears that there is a *prima facie* case to believe the answer is yes. Knowles *et al.*'s (2001) results, for example, imply that the stop rates for innocent blacks must exceed whites given the oversampling of blacks unless $\Pr(I|B) > \Pr(I|W)$, a condition that is implausible given the general tendency of crime rates to be higher among blacks than whites; this condition is of course inconsistent with the standard rationale for profiling made by its advocates. I am unaware of any evidence that (8) holds under profiling as it has been practised nor am I aware of anyone who has argued that the police are better able to identify black criminals than white ones when stopping automobiles, so that the profiling does not place a disproportionate burden on innocent blacks. For Knowles *et al.*, $\Pr(I|B)$ would have to be over twice as large as $\Pr(I|W)$ in order for the profiling pattern they studied to be consistent with (8), which I believe can be ruled out as implausible.

Thacher (2002) makes a very similar argument to the one presented here. Following Dworkin (2000), he argues that 'morally homogeneous' (Thacher, 2002, p. 8) groups should be stopped and searched at the same rate as violations of this would violate the

principle that a government should 'exhibit equal concern for each citizen's liberty' (Thacher, 2002, p. 10). This leads him to argue that stop and search policies should obey (8) as well as

$$\Pr(S|G, W) = \Pr(S|G, B), \quad (12)$$

which equates stop and search probabilities among the guilty as well, and by implication makes equal stop rates across races the only fair policy.

In my view, (12) is not necessary for fairness in the same way that (8) is. The reason for this is that (8) ensures fairness for individuals who have made decisions which society considers proper, whereas (12) does not. An individual can achieve equal treatment if (8) holds so long as he chooses not to commit crimes. This does not mean that violations of (12) cannot occur for unjust reasons; my claim is rather a violation of (12) is not by itself determinative of unfairness in the same way that (8) is determinative. If (12) is violated because there is some exogenous reason why it is easier to detect guilty members of one group versus another (for example, because the poorer group invests less in efforts to avoid detection), then violations of (12) do not seem offensive. This is the corollary to the argument made above that fair stop and search policies need not equalise stop rates across races, as discussed in reference to (10); differential rates are compatible with (8) for certain differences in the accuracy of detecting criminals.

To be clear, it is possible that (12) reflects unfairness. If (12) is violated because of a disproportionate interest in identifying guilty blacks, then fairness is a consideration that needs to be addressed. So, if the police invest in technology that allows identification of black criminals when investment in a different technology that allows identification of white criminals has been rejected for reasons unrelated to minimising the overall crime rate, e.g. prejudice, then one can construct objections to the violation.

Put differently, the importance of fairness may be reasonably linked to individual responsibility. An individual's claims to fair treatment by a government can be conditioned on certain requirements for individual behaviour, in this case, law abiding behaviour. To see why arguments that ignore this can go awry, consider the issue of selective prosecution of war criminals. Browning (1992) studies a particular reserve army unit, Reserve Police Battalion 101, which was involved in civilian killings in World War II. Because the members of the battalion were primarily from the same city (Hamburg), the records for the battalion were unusually complete, so the German government was able to prosecute members of the battalion. Prosecutions of war criminals of this type were quite rare (Browning, 1992, p. 146). Does the fact that one group of soldiers was singled out for prosecution because of a factor that was not under their control (the fact that they happened to end up in this particular unit) mean that they should not have been prosecuted? Clearly not, since their actions determined their membership in the group, i.e. the decision to participate in war crimes.

This issue of illegitimate choice and claims to fairness leads to a final issue that needs to be addressed. To what extent should an individual be held responsible for his choices? This question is less trivial than it appears, as may be seen in Roemer's (1993) example of indemnification of individuals for the medical costs induced by smoking. Consider groups of smokers where the group is defined by characteristics that are not a choice variable, such as ethnicity. Roemer argues that if a majority of individuals in a

group smoke, then society has an obligation to share medical expenses with the smokers in a way that it would not for a group where smoking is an outlier behaviour. Roemer's idea is that the median behaviour in a group reflects the factors to which group members have been exposed (e.g. role models, social norms etc.) for which they cannot be held responsible.

I claim that whatever the strength of Roemer's approach, it does not naturally extrapolate to issues of the treatment of those who commit crimes. In contrast, consider the question of war crimes guilt. One of Browning's findings is that 80–90% of the soldiers in Order Battalion 101 committed the war crimes when ordered to do so, and did so in the knowledge that they would not be punished for refusing. Does the fact that a majority engaged in the action have the same salience in terms of the appropriate levels of condemnation and punishment? While one may see an argument for some mitigation (though I do not), there is a key difference between this case and Roemer's example of cigarette smoking, one that suggests that mitigation should be weaker here than for Roemer. The difference is that harm of cigarette smoking accrues to the smoker whereas the acts of members of the order battalion harmed others. So long as individuals are not automatons, society may reasonably expect them to overcome social pressure and influences to avoid clearly immoral acts. This matters for my argument as I have differentiated the innocent and guilty in terms of their claims to equally fair treatment by society; the argument I have made means their choice is morally relevant and thus allows one to differentiate between them. Of course, one cannot equate drug possession with war crimes; my point is to show that there is no abstract reason why my differential treatment of the innocent and guilty is incoherent. Further, I do not argue that there is no comparative aspect to justice, only that (8) and (12) have different status in fairness discussions.

The upshot of this discussion is that the levels of $\Pr(S|I,W)$ and $\Pr(S|I,B)$ are particularly important objects in assessing the costs of profiling on individuals who have claims on full consideration by society; *ceteris paribus*, lower values for these probabilities are always preferred. Fairness is violated whenever $\Pr(S|I,W) - \Pr(S|I,B) \neq 0$, which will allow for partial orderings of different values of this difference. And as argued above, as an empirical matter it is reasonable to conclude that current profiling practices violate fairness.

My conclusion on fairness is restrictive as it does not address the question of fairness in the context of the full range of government policies. As Risse and Zeckhauser argue, a differential burden on a group in one context may be countered by differential burdens on others in different contexts. Hence, one does not analyse the fairness of restricting a military draft to young men in absence of evaluating how the differential burden of taxes etc. is assigned. Risse and Zeckhauser (2004) refer to this as 'functioning reciprocity', whose key idea is that a given differential burden is justified if society has attempted to allocate all burdens as fairly as possible in light of the need for differential burdens to exist in particular contexts due to efficiency considerations. However, I do not see any reason to regard the differential burden placed on blacks via profiling in traffic stops as falling into this category given the status of blacks in American society. Risse and Zeckhauser appear to concur with this and therefore further argue that when overall burdens are unjustly distributed, the imposition of a particular unfair burden may be justified if it represents a net benefit to the group

bearing the disproportionate burden, i.e. blacks. For reasons delineated in my discussion of welfarist considerations, I conclude evidence for this view of the net effect of profiling on black welfare is weak. And of course, the lack of support for racial profiling among blacks makes any such argument difficult, as noted by Risse and Zeckhauser, as it requires a policy maker to assume a patronising stance with respect to this opposition.

4. Welfare, Fairness and the Analysis of Ambiguous Environments

One way to interpret a number of my arguments is that there is substantial model uncertainty present in the analysis of the effects of racial profiling. The available data are consistent with models in which the disincentive effects of profiling are large as well as with models in which the disincentive effects are nonexistent. Arguments about the costs of racial profiling at best rely on essentially anecdotal claims about the harm that is inflicted on African Americans by stops and searches. From this perspective, there is an interesting parallel between the assessment of racial profiling and recent attempts in economics to deal with 'ambiguous' economic environments, i.e. environments in which the true model of the economy is unknown and probabilities cannot be assigned to the possible models. Much of this work has focused on the case where the true model lies in some space of possible models but where one cannot assign probabilities to the elements of this space and thereby engages in standard Bayesian decision making under uncertainty.

One approach to resolving ambiguity is to adopt a rule for choosing a particular model to work within a model space. The best known approach of this type is the minimax approach where a policy maker assumes that the least favourable model among the elements of the model space is the 'true' model and chooses a policy in response to that; this strategy has been adopted in work on macroeconomic policy evaluation, cf. Hansen and Sargent (2006). The minimax approach sets a lower bound on the losses a policy maker will incur. It is often criticised as excessively risk averse; other criteria such as minimax regret (Manski, 2004), which argues in favour of decisions that minimise the cost of not possessing complete information, have been advocated. This amounts to another strategy for choosing a model in the model space and optimising against it.

The evaluation of racial profiling is very much an example where one must engage in decision making in an ambiguous environment as we have no basis for assigning probabilities to different potential values of the behavioural response probabilities $\Pr[G|W, \Pr(S|W)]$ and $\Pr[G|B, \Pr(S|B)]$, different levels of private and social costs to a stop and search etc. Can recent work on decision making under ambiguity be used to evaluate racial profiling?

For our purposes, it seems difficult to see how one could implement rules such as minimax or minimax regret to the profiling context. One reason for this is that the model space under which racial profiling must be characterised is not well defined. How does one incorporate potential stigma or abuse into the model space? How does one characterise the range of possible levels of emotional harm to profiling? Similar problems exist because of the absence of a clear preference ordering over the consequences of profiling. Even if the model space were clearly defined, in order to define the least favourable model or to define which actions minimise regret on the part of the decision maker one must specify how deterrence and fairness objectives should be

traded off. Preferences defined over the outcomes associated with racial profiling are not readily quantifiable in the way, say, the objective of minimising the weighted sum of the variance of output and inflation is for a monetary policy authority. Further, differences in preferences with respect to efficiency and ethical goals will render the reporting of an evaluation exercise problematic unless the full range of possible preference orderings is considered.

My analysis thus far suggests that there is an unambiguous ethical cost to profiling, violation of equal treatment of the innocent, which needs to be matched against the combined ambiguous deterrence effects and individual and social costs to stops. Assume that within the range of potential deterrence effects that it is possible that the deterrence effect is large enough to trump all other factors and justify profiling. Does the absence of precise information on the model space and associated probabilities of models being true mean that one simply has to conclude that assessments of profiling cannot come to a conclusion? I believe such a conclusion is overly pessimistic. As an alternative strategy for evaluating racial profiling given the ambiguous model uncertainty inherent in the problem, I propose an approach that represents a response to the nature of the competing objectives that are at the heart of the evaluation: ethics and efficiency in meeting law enforcement objectives.

In environments such as profiling, I propose that the ambiguity that attends the assessment of costs and benefits be adjudicated by a notion of *presumption*. By presumption, I mean that certain actions by the government should be presumed to be inappropriate unless an affirmative case is made in their favour. In particular, I would propose as an evaluative criterion what I term a Fairness Presumption:

A government policy that violates fairness in its treatment of individuals is presumed to be wrong and hence requires an affirmative defence. The burden of proof is on the advocate of the policy to argue that the violation meets other social goals in a way to overcome the violation.

The Fairness Presumption employs a number of terms whose content needs to be defined before it is operationalised. Most obviously, how does one characterise the burden of proof? In fact, I do not think explicit definition is necessary for purposes of assessing the notion of presumption in the abstract.¹⁸ What the definition requires is that an individual treat deviations from fairness as something that needs to be argued for. The purpose of the definition is not to resolve an issue, but rather to define a criterion by which one can structure a debate on the merits of a policy. Similarly, work on virtues of deliberate democracy, initiated by Habermas (1984, 1987) and nicely surveyed in the essays in Elster (1998) supports the notion that just decisions are those that derive from an appropriately structured process of interchange and debate. The presence of a presumption for fairness structures policy debate in a way that treats individual equality as a primary good. Dworkin (2000) makes an elaborate argument on the primacy of equality in ethical analysis; in the narrow context of this paper Dworkin's argument justifies the Fairness Presumption although Dworkin would go much farther.

¹⁸ Further, legal systems employ different standards for the burden of proof according to context. The required evidentiary threshold is higher in criminal than civil cases, for example.

This approach to thinking about racial profiling implicitly moves us away from conventional decision-theoretic modelling to the more nebulous world of political debate. As such, it is consistent with some trends in political philosophy. Scanlon (1998) for example, argues that one way to identify immoral actions is to identify those that cannot be reasonably justified. This sort of argument attempts to derive moral conclusions on the basis of how rational agents with reasonable value systems will adjudicate issues. The Fairness Presumption makes explicit a 'ground rule' for policy debate. The principle is a procedural one in that it means that for policy debates certain rules should exist to structure the resolution of disagreements. This has some relation to Hampshire (2000) who argues that justice in procedures is something on which a consensus may be formed on the basis of asking how disagreements may be rationally adjudicated and justice in substance, where disagreements are inevitable. Unlike Hampshire, I reject the clean division between procedural and substantive justice in that the rules of adjudication I propose embody substantive values. Unless the rules of adjudication embody some notion of what Nagel (1979, p. 111) calls the 'assumption of moral equality between persons', it is not clear that rational adjudication has any content; the value of my opportunity to make arguments presupposes that those sitting in judgment will care about them. As such, my approach addresses a criticism of Hampshire due to MacIntyre (2000).

The principle I have proposed gives a 'weak' priority to fairness in that other factors may overcome it. For example, one can imagine cases where an increase in unfairness is reasonably justified via appeal to Rawls' difference principle, i.e. the decrease in fairness is justified if it improves the situation of the worst off person in society. Similarly, a welfarist calculation can be used to justify an increase in expected average utility even though it violates fairness. Whether an increase in unfairness can be justified using such arguments is context-specific. The key to the principle is that the burden of proof is on the advocate of a policy if it decreases fairness.

Nothing said here solves the problem of model uncertainty as it pertains to racial profiling or any other policy questions. Rather, I propose what I believe is a reasonable rule to help resolve the indeterminacy on decision evaluation that occurs in ambiguous environments. In some ways this strategy is analogous to the approach taken by Bewley (2002) to adjudicate decision problems when preferences are incomplete. Bewley argues in favour of an 'inertia assumption' to resolve ambiguity; status quo behaviour is not altered unless there is a reason to do so. The Fairness Presumption can be interpreted in parallel as saying that one does not move from a more to less fair set of policies unless there is a reason to do so. To be clear, the inertia assumption and Fairness Presumption are substantively different principles; what they have in common is that each introduces an external factor into a decision problem that allows resolution of an ambiguous decision environment. It is not clear that one can do more than identify sensible principles when it is necessary to evaluate decisions in ambiguous cases.

Based on the Fairness Presumption, I conclude that the racial profiling in traffic stops and searches should be rejected as a law enforcement strategy. Nothing in the available empirical literature suggests, in my judgment, that the deterrence effects are plausibly large enough to meet what I regard as an appropriate level of burden of proof

to overcome the clear violation of fairness that occurs in that innocent blacks are stopped and searched more frequently than innocent whites.

5. Relation to Other Policies

In this Section I consider how the arguments of this article relate to analyses of other policies that involve conditioning on race: anti-terrorism profiling and affirmative action. My intent is not to explore either of these difficult issues in detail but rather to indicate links between my analysis of profiling and how one might analyse these questions.

5.1. *Anti-terrorist Policies*

One reason why discussions of racial profiling are topical concerns the role of profiling in the war on terrorism. Do my arguments against profiling in police stops and searches have force in anti-terrorism contexts? I believe that any analogies that may be drawn are weak, so that one can oppose profiling in traffic stops and consistently support some forms of profiling in anti-terrorist conflicts.

First, the cases fundamentally differ in terms of potential benefits. As I have emphasised above, assessment of racial profiling must be done in context in order to assess costs and benefits to individuals and in terms of tradeoffs between individual welfare and other social objectives. There is self-evidently a stronger case for profiling young Arab men stopped in vans in the vicinity of nuclear plants based upon potential harm relative to racial profiling in traffic stops and searches.

A second important distinction between the two traffic stop and terrorism cases lies in the extent to which profiling will be efficacious in affecting crime. One of my objections to racial profiling is the absence of any evidence that profiling in traffic stops is efficient in terms of combating crime. This latter argument was based on the claim that the only non-trivial effects of traffic stops on drug trafficking is via individual incentives, something for which we have no evidence. In contrast, in the case of nuclear plants, the individuals subjected to profiling will include a non-trivial part of the potential criminal group, so the issue is stopping the particular individuals rather than inducing disincentive effects for a larger population of potential criminals. The case that profiling will reduce the probability of a successful terrorist strike against a nuclear plant would seem to be much stronger.

Further, it is difficult to see how the harm to individuals who are investigated in the nuclear power plant case is likely to be high. The activity is easily avoided in a way that use of public highways is not. In addition, this narrow context makes it relatively unlikely that the group of young Arab men would suffer general stigmatisation if profiling is limited to very specific contexts such as this.

Generally, defences of profiling in traffic stops that employ analogies to terrorism fail because the costs and, especially, benefits are of different orders of magnitude. The fact that the press may be prohibited from publishing war plans has little bearing on other cases where the government attempts to exercise prior restraints on freedom of the press. Of course, none of this in any way justifies (or condemns) profiling of Arab Americans as is currently practised by law enforcement.

5.2. *Affirmative Action*

Affirmative action policies raise fairness issues that are closely related to those discussed in Section 3. After all, interpreted in its starkest form, affirmative action policies involve adding race as a conditioning variable for some outcome such as college admission or employment. The affirmative action context in fact is useful in clarifying the difficulties that exist in operationalising a concept such as the Fairness Presumption. To see this, consider the case of college admissions. Letting A denote admission, one might initially start with the notion that race should not have any bearing on admissions, so that

$$\Pr(A|B) = \Pr(A|W). \quad (13)$$

Clearly, this definition of fairness is inadequate as it fails to account for past academic achievement; this corresponds to the idea that admissions are a reward for accomplishment and so captures intuitions that surround the idea of merit. If we define past academic achievement as P , then one might wish to use a definition such as

$$\Pr(A|B, P) = \Pr(A|W, P). \quad (14)$$

However, once one thinks about past academic achievement, then additional problems arise. Suppose that a black student has attended an inferior school, something for which he is obviously not responsible. One might want to only condition on the component of past academic achievement that reflects effort E . But if we modify the definition so that

$$\Pr(A|B, E) = \Pr(A|W, E) \quad (15)$$

other problems arise. If stigma has affected a student's past effort, then the variable for which we hold an individual responsible is partly determined by a factor outside his control. The only point I wish to make here is that affirmative action is a morally complex problem, so much so that it is difficult to draw easy inferences as to what constitutes fairness. Roemer (1993, 1998) specifically argues that this is to be expected: disagreements about the appropriate conditioning factors in assessing fairness are inevitable. However, traffic stops do not seem to involve additional factors in the way that, say, college admissions and affirmative action do. So, I see no strong reason to believe that opposition to profiling traffic stops imposes any constraints on one's views concerning affirmative action.

6. Conclusions

My basic conclusion is simple: when assessed by welfarist and fairness arguments, the overall case in favour of racial profiling in traffic stops and searches is very weak. The main welfarist argument in support of profiling, that racial profiling reduces crime rates, has not been established empirically; studies of guilt rates and race do not identify the deterrence effects from the policies. Retribution benefits, at least in the context of traffic stops appear to be second-order in comparison to deterrence. In addition, there are no good reasons to believe that the harm of profiling to African

Americans is minor when considered either from the perspective of individual stops or the associated stigma that may be produced by a profiling policy. Moving from individual to social harm, I argue that profiling can be contributory to stigma against blacks. In contrast, racial profiling violates an appealing notion of fairness: the equal treatment of the innocent. This leads to a situation where, based on current information, one is trading an ill-defined (in a probability sense) degree of deterrence against a principle, namely fairness. I argue that a presumption should exist against implementing policies that violate fairness, so that the appropriate public policy conclusion is that profiling is unjustified.

One weakness of this discussion is the lack of attention to alternative policing strategies and crime patterns. To see how the first matters, part of Kennedy's (1997) objection to profiling is that the police can find alternative strategies for choosing who to stop and search, strategies that do not require conditioning on race. Kennedy does not provide specific discussion of alternative strategies and I am unaware of any author who does. Nevertheless, the case for profiling will be correspondingly weakened if Kennedy is correct. A proper analysis of profiling requires a full specification of the strategy set available to police, which suggests a useful area for research. Similarly, my discussion does not address the issue of opportunity cost with respect to police resources.

As for crime patterns, one can imagine that complicated issues of fairness arise when there are correlations between the race of an offender and the race of a victim.¹⁹ My analysis has assumed that the aggregate crime rate is a sufficient statistic for understanding the rate of individual victimisation, V . Suppose that black and white criminals tend to affect members of their own racial groups. If a policy maker is interested in fairness in the allocation of police resources from the perspective of equalisation of victimisation probabilities, i.e.,

$$\Pr(V|I, W) = \Pr(V|I, B) \quad (16)$$

then one could imagine a defence of profiling that derives from this. To be clear, there is no reason to believe (and Randall Kennedy's argument would militate against it) that profiling, as opposed to other policing strategies, is required for fairness with respect to victimisation.²⁰ But it is important to recognise the possibility that different notions of fairness (in this case equality in stops and searches of the innocent versus equality in victimisation of the innocent) may prove to conflict with better knowledge of the determinants of crime. I conjecture that the analysis of the Fairness Presumption in cases such as this would require that a hierarchy of fairness claims be developed in which one considers first the direct effects of policy on fairness (in this case the profiling strategy on the innocent) and then the indirect effects (in this case, effects of the strategy on crime rates) that reflect the differential knowledge available in assessing direct versus indirect effects. In other words, I conjecture that a hierarchical structure

¹⁹ I thank Petra Todd for stimulating this line of argument. Risse and Zeckhauser (2004) raise this issue in their argument that the crime deterrence benefits of profiling may disproportionately benefit blacks.

²⁰ I am not aware of any evidence that this is the case. In order to make such a claim, one would need to account for equilibrium changes in criminal behaviour, e.g. the locus of activity, in the presence of profiling. One could well imagine that traffic stops cause drug-related activities to concentrate more highly in the residential neighbourhoods of criminals.

to presumption exists; the burden of proof is on the proponent to show that a policy promotes fairness indirectly when its direct effects reduce fairness. Development of this argument is part of future work.

To conclude, one interpretation of this essay is that what is needed is better evidence on the effects of profiling strategies on individual decisions, which means either the construction of new data sets and/or econometric methods that allow a researcher to identify (at least partially, in Manski's (2003) sense) deterrence effects. Similarly, we have very little firm evidence on the effects of a stop and search on the well being of a motorist. Many of the arguments I have presented concern the plausibility of empirical claims that have been made by others or appear to be necessary to justify profiling and typically my conclusion has been that these claims are little more than assertions. For this reason, my conclusions should be read as contingent on our current ignorance.

University of Wisconsin

References

- Anwar, S. and Fang, H. (2004). 'An alternative test of racial prejudice in motor vehicle searches', mimeo, Department of Economics, Yale University.
- Berlin, I. (2002). 'Two concepts of liberty', in (H. Hardy, ed.), *Liberty*, Oxford: Oxford University Press.
- Bewley, T. (2002). 'Knightian decision theory, part I', *Decisions in Economics and Finance*, vol. 25(2), pp. 79–110.
- Browning, C. (1992). *Ordinary Men*. New York: Harper Collins.
- Cohen, G. (1989). 'On the currency of egalitarian justice', *Ethics*, vol. 99, pp. 906–44.
- Dominitz, J. (2003). 'How do the laws of probability constrain legislative and judicial efforts to stop racial profiling?', *American Law and Economics Journal*, vol. 5(2), pp. 412–32.
- Durlauf, S. (1999). 'The memberships theory of inequality: ideas and implications', in (E. Brezis and P. Temin, eds.), *Elites, Minorities, and Economic Growth*, pp. 161–77, Amsterdam: North Holland.
- Durlauf, S. (2002). 'The memberships theory of poverty: the role of group affiliations in determining socioeconomic outcomes', in (S. Danziger and R. Haveman, eds.), *Understanding Poverty in America*, pp. 392–416, Cambridge: Harvard University Press.
- Dworkin, R. (2000). *Sovereign Virtue*, Cambridge: Harvard University Press.
- Elster, J. (1998). *Deliberative Democracy*, New York: Cambridge University Press.
- Gross, S. and Barnes, K. (2002). 'Road work: racial profiling and drug interdiction on the highway'. *Michigan Law Review*, vol. 101(3), pp. 651–754.
- Habermas, J. (1984). *The Theory of Communicative Action*, vol. 1, Boston: Beacon Press.
- Habermas, J. (1987). *The Theory of Communicative Action*, vol. 2, Boston: Beacon Press.
- Hahn, F. (1982). 'On some difficulties of the utilitarian economist', in (A. Sen and B. Williams, eds.), *Utilitarianism and Beyond*, pp. 187–98, New York: Cambridge University Press.
- Hampshire, S. (2000). *Justice is Conflict*, Princeton: Princeton University Press.
- Hansen, L. and Sargent, T. (2006). *Robustness*, Princeton: Princeton University Press, forthcoming.
- 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*, vol. 71(4), pp. 1275–381.
- Kamm, F. (2000). 'Nonconsequentialism', in (H. LaFollette, ed.), *The Blackwell Guide to Ethical Theory*, pp. 205–26, Oxford: Blackwell Publishing.
- Kaplow, L. and Shavell, S. (2001). 'Any non-welfarist method of policy assessment violates the Pareto principle', *Journal of Political Economy*, vol. 109, pp. 281–6.
- Kennedy, R. (1997). *Race, Crime and the Law*, New York: Pantheon.
- Knowles, J., Persico, N., and Todd, P. (2001). 'Racial bias in motor vehicle searches: theory and evidence', *Journal of Political Economy*, vol. 109(1), pp. 203–29.
- Loury, G. (2002). *The Anatomy of Racial Inequality*, Cambridge: Harvard University Press.
- MacIntyre, A. (2000). 'Review of Justice is Conflict by Stuart Hampshire', *Mind*, vol. 109(435), pp. 618–21.
- Manski, C. (2003). *Partial Identification of Probability Distributions*, New York: Springer-Verlag.
- Manski, C. (2004). 'Statistical treatment rules for heterogeneous populations', *Econometrica*, vol. 72(4) pp. 1221–46.
- Nagel, T. (1979). 'Equality' in *Mortal Questions*, Cambridge: Cambridge University Press.

- Persico, N. (2002). 'Racial profiling, fairness, and the effectiveness of policing', *American Economic Review*, vol. 92(5), pp. 1472–97.
- Petit, P. (1991). 'Consequentialism', in (P. Singer, ed.), *A Companion to Ethics*, pp. 230–40, Oxford: Blackwell Publishing.
- Risse, M. and Zeckhauser, R. (2004). 'Racial profiling', *Philosophy and Public Affairs*, vol. 32 (2), pp. 131–70.
- Roemer, J. (1993). 'A pragmatic theory of responsibility for the egalitarian planner', *Philosophy and Public Affairs*, vol. 10, pp. 146–66.
- Roemer, J. (1998). *Equality of Opportunity*, Harvard: Harvard University Press.
- Rogers, R. and Prentice-Dunn, S. (1981). 'Deindividuation and anger-mediated interracial aggression: unmasking regressive racism', *Journal of Personality and Social Psychology*, vol. 41(1), pp. 63–7.
- Scanlon, T. (1998). *What We Owe To Each Other*, Cambridge: Harvard University Press.
- Sen, A. (1970). 'The impossibility of a Paretian liberal', *Journal of Political Economy*, vol. 78, pp. 152–7.
- Sen, A. (1979). 'Utilitarianism and welfarism', *Journal of Philosophy*, vol. 76(9), pp. 463–89.
- Sherif, M., Harvey, O., White, B., Hood, W. and Sherif, C. (1961). *Intergroup Conflict and Cooperation: The Robbers Cave Experiment*, Norman: Institute of Group Relations, University of Oklahoma, reprinted by Wesleyan University Press, 1988.
- Steele, C. (1997). 'A threat in the air: how stereotypes shape the intellectual identities and performances of African Americans', *American Psychologist*, vol. 52, pp. 613–29.
- Thacher, D. (2002). 'From racial profiling to racial inequality: rethinking equity in police stops and searches', mimeo, School of Public Policy, University of Michigan.