

THE IMPOSSIBILITY OF PARTICULARISM

Frederick Schauer¹

In an era pervaded by concerns about terrorism, the topic of ethnic and racial profiling has often taken central stage. If certain forms of unlawful violence are disproportionately committed by people of one race, religion, or ethnicity, as it is often argued, then it is only rational to target members of that race, religion, or ethnicity in an attempt to prevent and to prosecute terrorist and related violent acts. But such targeting is profoundly unjust, so the equally common response goes, for morality demands that people be treated as individuals and not as members of groups. To encumber people with the non-universal characteristics of the groups of which they may be members, it is said, is to deprive individuals of their identity and of their agency, and is also to unfairly impose on people characteristics they may not share, and behavioral inclinations they may not possess.

The debate about racial, religious, and ethnic profiling is important in its own right, but it is also an instantiation of a larger debate about generalization and particularity in decision-making more generally. Should decisions be made on the basis of probabilistic but non-universal generalizations, as is also often argued, or is it morally necessary, even outside the

¹Frank Stanton Professor of the First Amendment, John F. Kennedy School of Government, Harvard University. This paper has been prepared for Legal Theory Workshop at the Columbia Law School. An earlier version was presented at the Conference on Discrimination Law and Policy in Contemporary Europe held on 26-27 May, 2006, at the Robert Schuman Center for Advanced Studies of the European University Institute, Florence, Italy. I am happy to acknowledge research support from the Joan Shorenstein Center on the Press, Politics and Public Policy, Harvard University.

domain of debates about profiling, to treat every person as an individual and to understand every decision-making occasion in its full contextual richness. Moreover, the debates about profiling and discrimination connect with contemporary philosophical debates about so-called moral particularism.² Is it in the very nature of morality, as Jonathan Dancy and others maintain, that moral judgments can only attach to particular actions and not to larger classes of actions, or is such a claim “almost incoherent,” as is alleged by philosophers such as Walter Sinnott-Armstrong?³

My goal in this paper is to address these larger debates about generalization and particularization, and also to consider the moral ontology lying behind the typical particularist arguments. For if it is the case, as I shall argue, that a pure particularism is impossible, then we can understand the debates about particularism and their application to issues such as racial, religious, and ethnic profiling in a different and decidedly cooler light.

I

The distinction between the general and the particular is of course highly familiar and often analyzed, but it is not my intention here to engage the deepest questions of philosophical

²See, e.g., Jonathan Dancy, Moral Reasons (Oxford: Blackwell, 1993); Brad Hooker & Margaret Little, eds., Moral Particularism (Oxford: Oxford University Press, 2000), John McDowell, Mind, Value, and Reality (Cambridge, Massachusetts: Harvard University Press, 1998); David McNaughton, Moral Vision (Oxford: Blackwell, 1988). A related form of particularism features prominently in some of feminist theory. See Mary Field Belenky, et al., Women’s Ways of Knowing: The Development of Self, Voice, and Mind (New ork: Basic books, 1997); Katharine T. Bartlett, “Feminist Legal Methods,” Harvard Law Review, vol. 103 (1990), pp. 849 ff.; Margaret Walker, “Moral Understandings: Alternative ‘Epistemologies’ for a Feminist Ethics,” Hypatia, vol. 4 (1989), pp. 15-28. And a more tempered form of particularism can be found in works of legal and constitutional theory such as Cass R. Sunstein, One Case at a Time: Judicial Minimalism on the Supreme Court (Cambridge, Massachusetts: Harvard University Press, 1999).

³Walter Sinnott-Armstrong, “Some Varieties of Particularism,” Metaphilosophy, vol. 30 (1999), pp. 1-12.

logic. Rather, and to bring the discussion at least slightly down to earth, I want to begin with a simple model. So let us assume that we have ten individuals, and let us assume further that each of those individuals has ten characteristics. And then suppose that we give each of those characteristics a value. So one characteristic would be eye color, with values of blue, green, gray, and brown. Another would be hair color, and there the values would be blond, brown, black, and red. Still another would be religion, with values of, say, Catholic, Protestant, Muslim, Buddhist, Jewish, and atheist. And so too for gender, race, national origin, age, height, educational level, and occupation.

In this simplified model, each of the ten individuals in the population can thus be described in terms of the ten characteristics. One of the individuals will be, say, a blue-eyed, red-haired, Caucasian Australian Muslim woman who is thirty-seven years old, stands five feet seven inches tall, has a bachelor's degree, and is employed as an accountant. Analogous descriptions can also be given for the other nine. So now suppose that for each of the characteristics a particular value is shared by two or more individuals, but that no two individuals have identical values for all ten of the characteristics.

With this extremely simplified example in hand, we can understand the commonplace, even if not the formal, distinction between the general and the particular. Because no two individuals in our hypothetical array share identical values for all ten characteristics, each individual has a unique pattern of characteristics, and thus the ten individuals are all in some ways different. But because for each of the characteristics a particular value is shared by two or more individuals, the category of Muslims will be, say, a category of three people, the category of women will be a category of five, the category of the green-eyed will be a category of two, the category of the brown-haired will be a category of four, and so on. And thus although each of the individuals in the array has a different set of characteristics than all other members of the array, there is no one characteristic that is not shared with anyone. And as a result of this

distribution of characteristics, the process of generalization can be seen as one involving making a decision on the basis of one or more of these shared characteristics, but the process of particularization would involve making decisions based only on a combination of all of the characteristics, a combination that we have stipulated, at least for purposes of this example, to be unique to each of the ten individuals.

II

In this example, a decision procedure employing all ten characteristics will appear to be a particularistic one, but things are not so simple. Each of the characteristics in the previous example is itself a generalization, and, without getting too deeply into contested questions of metaphysics, we can say that to identify someone as Muslim, or as an accountant, or as green-eyed, for example, itself involves a process of generalization. The shade of a person's eyes occupies an exact point on a spectrum, but the part of the spectrum we label "green" is more than one point, and is in fact a range. As a result, to have an eye color that we label "green" is to have an eye color located within a range on the spectrum, but not necessarily located at the same point on the range as other people whose eyes will also be designated as green. And so too with the array of characteristics that would make someone a Muslim, an accountant, or anything else. All of the individuals who are entitled to be called "accountants" differ, and they differ even with respect to the characteristics of accounting itself. Similarly, some of the people who are entitled to call themselves lawyers or less lawyerly (whatever that means, but that is not the point) than others, but the category of lawyers, like the category of accountants and the category of Muslims, groups together those who are different even with respect to the very criteria for inclusion within the category.⁴

Insofar as the primary building blocks of our perception are sense data, therefore, even

⁴[Footnote on Rawls and "range" attributes and categories. And on Waldron's use of this idea].

the particular characteristics of particular individuals are groupings of sense data, and the labels we apply to certain groups are applied to something broader than a unique array of sense data. And from this perspective what looks like a particular characteristic is in fact a generalization. Indeed, even the more extreme examples – fingerprints, prototypically, or, even more extreme, DNA patterns – exhibit the same phenomenon, for although the likelihood of two individuals sharing the same fingerprint, or, even less likely, the same DNA pattern is small, it is not impossible. As a result, to identify someone as having such-and-such DNA or such-and-such fingerprint pattern is not different in kind, although it is different in degree, from identifying them as being members of the class of people with that DNA or that fingerprint pattern, even as we recognize that the categories in cases like fingerprints or DNA are very likely, but not inevitably, to be categories with one member.

What we call particularism, therefore, is also not dramatically different in kind from what we call generalization. Rather, to focus on the particular characteristics of an individual is to take into account a larger rather than smaller number of characteristics, or attributes, with the consequence being that it is more likely that the set of people with the same array of attributes will be extremely small. But this is a quantitative difference only (which is not to say it is unimportant or non-existent), and thus the call to treat people as individuals is best understood as a call to take into account more rather than fewer of their attributes, even as we recognize that each of those attributes is shared with many others, and even as we recognize that a certain array of attributes is likely to be shared with fewer others, but not necessarily unique in either a logical or an empirical sense.

III

Once we understand that even the most particularized particularization is based on a series of generalizations, we can see that the distinction between generalized and particularized decision-making is largely one of degree. That the distinction is scalar does not mean that

getting it right is not important, morally or politically or epistemically. But it does mean that there is not, popular rhetoric notwithstanding, anything immoral about generalization tout court, for generalization is inevitable. And because generalization is inevitable, so too is discrimination. That is, when we make decisions about people, or that have an effect on people, we are using some of their generalized characteristics to make probabilistic assessments about their past, present, or future behavior. Or we are using some of their available characteristics or attributes to make assessments of other characteristics or attributes that may be less available. With some frequency, we refer to the use of an available fact to determine the existence (or not) of some unavailable fact as a proxy.

Some of this may not be obvious, so it is worth lingering over it. And let us take our previous stylized example. So if we wanted to predict who was likely to commit an act of terrorism at a railway station (or who was the person most likely to already have committed an act of terrorism at a railway station), we might, on the basis of past data, predict that the person with a particular value for each of the ten characteristics was the one who was most likely to engage in the terrorist act, and then to take steps accordingly. Now of course some of these values will be spurious (which is not to say that a different value for the same characteristic will necessarily be spurious). There is no evidence, for example, that hair or eye color, or day of birth, or any of a large number of other attributes are at all predictive of (almost) anything, let alone terrorism. So were we to do a multiple regression with all values for all characteristics and for all behaviors of concern (of which terrorism is but one), we could derive, in theory and indeed in practice (as is well known to the tax authorities), an algorithm of all of the non-spurious predictors in proportion to their predictive contribution to assessing that which is in fact our primary concern.

One conclusion that emerges from this exercise is the identification of one form of discrimination. And that form is the use for predictive purposes of factors that simply do not

predict, or the use of epistemic proxies that are of no epistemic value at all. This form of discrimination, the use of non-predictive predictors and the use of statistically spurious proxies, is little more than superstition, as historically common as it may be, and it is hard to understand any plausible justification for it. So consider, for example, the belief in some parts of the world, even today, that blood type is a non-spurious predictor of various behavioral characteristics. There is, of course, no evidence whatsoever for this proposition, any more than there is evidence that the predictions implicit in astrological forecasts or phrenological evaluations have any empirical basis. So although it may be widely believed at certain places and at certain times that having AB-positive blood is predictive of, indicative of, or causal of, hard work, that belief is simply false. And the universe of false beliefs of this variety is hardly so restricted. One form of discrimination, therefore, and a form properly condemned, is the practice of distinguishing among people for some valid purpose (it is not a bad thing to distinguish the terrorists from others, for example) but using as the discriminating factor some characteristic that is in fact a statistical failure as a discriminator.

IV

This is an appropriate place to emphasize that the assessment of failure of some discriminating factor is decidedly not a function of that factor being an imperfect predictor or imprecise proxy. If people with green eyes were more likely to be terrorists than were randomly selected members of the population at large (or than were the members of some relevant comparison group), then possessing green eyes would be predictive of terrorism, even if most people with green eyes never committed terrorist acts, and even if most people who committed terrorist acts did not have green eyes. To take an example from a contemporary controversy in many parts of the world, a dog being a pit bull is predictive of that dog being dangerous, even if the vast majority of pit bulls are not dangerous and even if the vast majority of dangerous dogs are not pit bulls.⁵ As long as the attribute of being a pit bull increases the

⁵See Frederick Schauer, Profiles, Probabilities, and Stereotypes (Cambridge, Massachusetts:

accuracy of a prediction of some behavior over what it would be for the same population (all dogs, for example) but without using the attribute, the attribute is properly understood to be predictive, and to be a non-spurious proxy for, in this case, dangerousness. And because an attribute's being predictive in this sense is highly variable from slightly predictive to very predictive, the word I use to capture the range, and to include within the range anything that is at least slightly predictive, is non-spurious. A spurious attribute helps us not at all in predicting behavior or in determining some unseen fact, and thus a non-spurious attribute is any attribute that makes at least some contribution to the predictive or the epistemic enterprise. Obviously some predictors and proxies are better than others, and most very good predictors and proxies are non-universal, being both under- and over-inclusive. Still, the first important distinction is between the spurious and the non-spurious, a distinction that enables us properly to condemn the use of spurious proxies while at the same time understanding that virtually all proxies are imperfect and that the use of imperfect proxies in decision-making is inevitable.

Although I often use the word "prediction," I repeat that foregoing is not limited to the prediction of future behavior, and includes assessment of past acts or events, and also the identification of unknown facts. If we want to know y , and if knowing x helps us to know y , then x is non-spurious even if y is a physical fact, or is an event that has already taken place. In that sense the best description of a non-spurious x is that it is indicative, where the presence of x can help indicate something less known, regardless of whether that something is an existing fact, a past event, or the probability of some future behavior.

It should also be clear from the foregoing that the relationship I am describing is an epistemic one, and need not have a causal component. Smoke does not cause fire, but that does not detract from the fact that the presence of smoke is a good indicator of fire. And although

fire does cause smoke, there is no causation either from the indicator to the indicated or from the indicated to the indicator in the case of the relationship between baldness and sound judgment, but if baldness correlates with age and age correlates (within some range) with sound judgment, then baldness will be a non-spurious indicator of the existence of sound judgment. Similarly, if being from the Netherlands is indicative of being a bad driver (which it almost certainly is not), and if being from the Netherlands is indicative of driving an orange automobile (which it may well be, because orange is the national color of the country), then possession of orange car would be a non-spurious indicator of poor driving, even though there was no causal relationship between the color of the car and the skill of the driver, or between the skill of the driver and the color of the car.

V

It should be apparent that using the word “discrimination” is unlikely to be helpful to analyzing the issues. Once we see that even maximal particularism involves at least some categorization and generalization, we can see as well that all non-spurious indicators will discriminate. Having identified the relevant trait – terrorism, dangerousness, financial irresponsibility (or responsibility), or whatever – we then use non-spurious indicators to discriminate between those who are more likely to have the trait from those who are less likely to have it. Governments discriminate on the basis of non-spurious indicators when they set minimum ages for voting and drinking and driving and maximum ages for piloting commercial airplanes, just as they discriminate when they prohibit convicted child molesters from being employed as school teachers.

As discussed previously, we might choose to limit the word “discriminate,” with its increasingly pejorative connotations, to those forms of discrimination that are based on spurious indicators. But this usage is non-standard and would replace one problem with another. So it seems therefore preferable to recognize that all predictors and all proxies (and all laws, for that

matter) will discriminate, but that there is benign discrimination and pernicious discrimination. Thus, we will have the tools to recognize that not all discrimination is pernicious. But although it should be clear by now that there are many forms of benign – indeed, inevitable – discrimination that are based on non-spurious but nevertheless imperfect proxies, it does not follow from this that all uses of non-spurious proxies are benign. This last point is both different and important, and thus it is to the domain of the non-spurious but nevertheless problematic proxies that I now turn.

VI

Once we recognize that discrimination is inevitable, that most of it equally inevitably involves the use of imperfect proxies, and that pure particularism is impossible, the debate is then revealed as being about how particular we should be. Although we persistently make decisions on the basis of probabilistic – non-spurious – but non-universal indicators, it is still open to us to use more or less imperfect indicators, recognizing the less imperfect indicators will still be somewhat imperfect, and that the least imperfect indicators may be quite costly in a number of ways. Or, to put it differently, once we understand that virtually all decisions involve the use of proxies, one question is how broad should those proxies be, and another is when is it permissible to employ less rather than more accurate proxies.

A common problem is the use of non-spurious but less accurate proxies in place of more accurate ones, and to do so for any of a number of reasons, some good and some less so. So if it turns out that ethnicity of some kind (or race, or national origin) is indeed predictive of terrorism, then the best proxy for purposes of, say, airport passenger and luggage screening, will be an algorithm using all of the factors (appropriately weighted), including ethnicity, that predict terrorism. But suppose (increasingly counterfactually, in the computer era) that the cost of employing a 47-factor algorithm is simply too high, regardless of whether that cost is measured in terms of time to apply it, or cost of training enforcers in its use, or something else. But

regardless of the nature of the cost, it will often be the case that the cost of using the best multiple regression-generated algorithm will be prohibitive, in which case the task will become one of selecting a smaller number of non-spurious factors from among the 47 non-spurious factors in order to produce a workable, even if less accurate in theory, algorithm.

Once we move from the best algorithm to the best practically usable algorithm, there is a substantial risk, as we have far too often seen in practice, that ethnicity, race, or religion will be employed as among the small number of components in a simplified algorithm. The risk comes not from the basic idea of factors such as these being part of the simplified algorithm, but rather from the selection of such factors not because they are among the three or four most predictive, and not because an algorithm employing them is the most predictive of all of the possible three- or four-factor algorithms, and not because the visibility of such factors makes them most efficient to use, but because these factors are more salient independent of their relative reliability, or because they are believed to be more predictive than they actually are.

Thus, it would not be implausible to imagine that the best full algorithm for predicting terrorism would be something like the hypothetical 47-factor algorithm just noted, and would be one including, say, ethnicity, or national origin. And it is also plausible to suppose that the best seven-factor algorithm -- the best algorithm that can be constructed using no more than seven factors from the 47-factor full list -- would be an algorithm seeking to identify those who are (1) male; (2) between the ages of 19 and 30; (3) unemployed; (4) possessing no credit card and paying for their tickets in cash; (5) not checking luggage and having only carry-ons; (6) not members of a frequent flier program; and (7) having Middle Eastern ethnicity and/or national origin. And perhaps, to continue the example, the best three-factor algorithm would be one seeking to identify (1) unemployed (2) males (3) between the ages of 19 and 30. Were this to be so, there might nevertheless be a substantial risk that those who constructed the algorithm (which might exist informally in the heads of airport personnel just as much as it exists in formal

rules, regulations, and instructions), if required to choose a three-factor algorithm, would choose (1) Middle Eastern (2) males (3) between the ages of 19 and 30 rather than (1) unemployed (2) males (3) between the ages of 19 and 30, and would do so despite the fact that being unemployed was more predictive than being Middle Eastern, even though being Middle Eastern was not, by hypothesis, spurious.

We can imagine many reasons for such selection of less predictive rather than more predictive characteristics. Some would be simple racism. Others would be based on erroneous beliefs that racial and ethnic and religious, say, factors are more predictive than they really are, even though in some circumstances that may be somewhat predictive. And still others would be based on visibility or other forms of salience or availability. People who get nervous when others who are likely Arab Muslims sit next to them on the airplane are less likely to get nervous when others who are convicted violent felons sit next to them, and that is because identifying someone as an Arab Muslim is easier – more visible, and thus more salient, or more available – than identifying someone as a convicted felon. And this is so even if, as is in fact the case, the fact of someone being a convicted felon is far more predictive of something dangerous – a better proxy – than is the fact of someone being an Arab Muslim.

VII

Once we recognize the problem, we can understand why the non-spuriousness of some factor may not be a sufficient condition for its use as a proxy. For the very same factors that would lead people to exaggerate the relative importance of non-spurious indicators would likely lead them to overvalue the predictive contribution of those factors as well. As a result, a strong argument for disallowing the use of non-spurious factors is that they are so likely to be overused that we mandate their underuse. Sometimes the likelihood of overuse comes from salience or availability. Sometimes it will come from erroneous beliefs about the accuracy or reliability of those non-spurious factors. And sometimes it will come from the view that there are some

values that may be more important than accuracy, or will at least exist alongside accuracy as important values.

VIII

Once we understand that pure particularism is impossible and that the use of non-spurious but non-universal proxies are an inevitable part of thinking, talking, and deciding, and once we understand as well that the wrong proxies (even the wrong non-spurious ones) may often be selected for inappropriate reasons, the question still remains about the appropriate size of the proxy. To return to the stylized example I used earlier, what should guide the choice of how many of the ten attributes to use in making a decision or a prediction, even while recognizing that using all ten still involves using generalizations?

If it is taken as a given that each of the ten attributes are non-spurious for some decision, it follows that using more than fewer attributes will increase decisional accuracy, assuming (probably incorrectly) for the sake of argument that increasing the number of attributes used will not increase the error rate in applying those attributes. But accuracy of course comes at some cost, and thus it would not be difficult to imagine, even if not actually to calculate, a consequentialist calculation leading to a consequentially optimal degree of particularization for some decision.

At this point there is a familiar and a non-familiar qualification. The familiar qualification is that there may still be non-consequentialist reasons for excluding even some accuracy-enhancing attributes from the portfolio of attributes that are used to make the decision. Even apart from the possibility that attributes of race, ethnicity, national origin, and the like may because of their cognitive availability or their special likelihood of misuse in fact decrease aggregate accuracy, using them may still implicate any of a number of by-now familiar non-consequentialist side-constraints on consequentialist, utilitarian, welfare, utility, or accuracy

maximization.

The less familiar qualification, however, is the question whether, questions of accuracy aside, and questions of morally (or legally) problematic attributes aside, there is within the concept of justice itself a requirement of maximum particularization. Still assuming that every eligible attribute is non-spurious, and still assuming (counterfactually) that increasing the number of attributes used will not decrease accuracy, then the failure to use the maximum number of attributes will simply be irrational if increasing the number of attributes comes at no cost. But if increasing the number of attributes used does increase the cost, then the question is whether there are justice-based side-constraints requiring the use of more attributes even when doing so increases the cost of the decision, or whether (and this is simply saying the same thing in a somewhat different way) there is some justice-based moral residue when for the sake of efficiency (or even, relaxing one of the assumptions above, for the sake of accuracy) we use something less than the maximum number of non-spurious attributes. When we make decisions on the basis of a smaller rather than larger number of non-spurious attributes (or generalizations), and when the attributes used or ignored do not themselves as attributes carry moral baggage, is there an element of injustice involved in failing to particularize to the maximal degree? In other words, is particularization itself a moral requirement, independent of the actual character of the attributes employed or ignored?

Much of our existing discourse about “treating everyone as an individual” appears to assume a positive answer to the question that ended the preceding paragraph, but it is not clear that there is an existing argument producing that conclusion. There are arguments relying heavily on the particular characteristics of attributes such as race, ethnicity, gender, religion, sexual orientation, age, national origin, and the like, but when all of these factors are eliminated, and the question is stripped down to the question of whether particularism qua particularism is a component of justice and a mandate of morality such that it would constrain (even if not

absolutely) the consequentially-justified failure to pursue the maximum non-spurious particularization, it is by no means clear that such a requirement exists, and what the argument for its existence would look like.

Thus, even if a full particularism is impossible, and thus even if making decisions on the basis of non-universal indicators is an inevitable part of all decision-making, moral and otherwise, there is the question whether using a combination of more rather than fewer indicators is itself a requirement of morality. And although much more could be said here, it may be worth noting that large and imperfect proxies may not only serve values of efficiency and cost-saving, but may also serve the moral values of community. When, to take a highly controversial example, the food safety regulators of the European Union (the bureaucrats in Brussels, as it is so often put) treat the cheese-makers of Spain, Italy, and France in the same way as they treat the cheese-makers from Germany and the Netherlands, they may be ignoring the best proxies for food safety and other valuable ends. But in doing so they are generalizing, and treating unlike cases alike, and thus creating the very idea of a community. Communities are created by, perhaps even defined by, the use of shared characteristics from among a much larger array of available characteristics. “All men are created equal,” as it is put in the American Declaration of Independence, captures this well, since of course all men (sic) are decidedly not created equal. To say that all men, or all cheese-makers, or all dogs, or all whatever, are equal, is ascriptive and not descriptive, and represents a commitment to use some characteristics as the relevant discriminators in the face of relevant distinctions among those who are grouped together. We generalize, therefore, not only to save time, and not only because it is inevitable, but because generalization, rather than keeping us apart, is what brings us together.