

## IT'S NOT THE THOUGHT THAT COUNTS

Deborah Hellman\*

Draft: June 1, 2005

### Abstract

The article considers a central question about discrimination – are an actor’s intentions relevant to whether an action wrongfully discriminates – and takes issue with a familiar answer to this question. If one thinks of “discrimination” in its literal sense, as simply drawing distinctions among people on the basis of possessing or lacking some trait, it becomes clear that discrimination is ubiquitous and often benign. The challenge is to distinguish when discrimination is permissible and when it is not. One common answer to this question is that it is the intentions of the actor who adopts or enacts a law, policy or decision that are crucial. Legal doctrine, both constitutional and statutory, reflects this view by treating the actor’s intentions as centrally important. But is the moral claim on which it rests defensible; are intentions morally relevant to whether discrimination is wrong?

The article argues that the actor’s intent in enacting a law, adopting a policy or making a decision is irrelevant to the moral assessment of whether the law, policy or decision wrongfully discriminates. The article begins in Part I by drawing an analogy to a debate in the philosophical literature about the Doctrine of Double Effect in order to press the point that the focus on intentions confuses assessment of the wrongfulness of the action with assessment of the moral blameworthiness of the actor. The article goes on to argue that when we look more closely at instances of discrimination, we see that it is not the aims of the actor that render the action wrongful, rather it is what the actor does, whether intentional or unintentional, that matters. Parts II and III raise and reply to objections to the arguments advanced in Part I. In particular, Part II explores the argument that the actor’s intent is necessary to determine whether or not a law or policy distinguishes among people on the basis of a suspect trait. Part III explores objections meant to show that bad intentions contribute, at the very least, to rendering an action wrongful. The article finishes by concluding that, as far as discrimination is concerned, it is not the thought that counts.

---

\* Professor of Law, University of Maryland School of Law (on leave 2004-05); Eugene P. Beard Faculty Fellow in Ethics, The Edmond J. Safra Foundation Center for Ethics, Harvard University. I would like to thank Larry Alexander, Mitchell Berman and Richard Fallon for their detailed comments on earlier drafts of this article. In addition, I would like to thank the participants in the Legal Theory Workshops at UCLA and Dartmouth College for very helpful feedback, especially Seana Shiffrin and Julia Driver. Finally, I would like to thank the fellows at the Center for Ethics for their company and critique while I worked on this and other projects throughout the year. Megan Moran-Gates provided very helpful research assistance.

## Table of Contents

|                                                                                     |    |
|-------------------------------------------------------------------------------------|----|
| Introduction                                                                        | 2  |
| Part I: The Case Against Intention                                                  | 7  |
| A. Intention and Identification                                                     | 8  |
| B. Intention and Evaluation                                                         | 11 |
| Part II: Is Intention Necessary to <i>Identify</i> the Classifying Trait?           | 20 |
| A. Intention Is Neither Necessary Nor Sufficient                                    | 22 |
| B. Facial Discrimination, Disparate Impact and Subjective Intent                    | 26 |
| C. Private Bias                                                                     | 29 |
| Part III: Is Invidious Intent Relevant to the <i>Evaluation</i> of Classifications? | 33 |
| A. The Target                                                                       | 34 |
| B. The Purpose                                                                      | 42 |
| Conclusion                                                                          | 47 |

### Introduction

In an oft-quoted passage, Justice Holmes asserts that “even a dog distinguishes between being stumbled over and being kicked.”<sup>1</sup> According to Fred Schauer, Holmes overestimates dogs.<sup>2</sup> Perhaps Holmes is wrong about people too. The passage appears to suggest that it is the intention of the actor that determines what sort of action takes place. The actor’s intention has the power to transform the action from an instance of kicking to an instance of tripping. And because we assume that kicking is, generally speaking, morally wrong while tripping is, generally speaking, not morally wrong, this transformation is important. But this common understanding of the meaning of that passage<sup>3</sup> may be mistaken. After all, Holmes only needs to bother with the assertion that

---

<sup>1</sup> OLIVER WENDELL HOLMES, JR., *THE COMMON LAW* 7 (Transaction Publishers, 2005).

<sup>2</sup> See Frederick Schauer, *Intentions, Conventions and the First Amendment: The Case of Cross-Burning*, 6 SUPREME CT. REV. 197, 197 (2003) (asserting that “[i]n claiming that ‘even a dog distinguishes between being stumbled over and being kicked’ Justice Holmes demonstrated his limited knowledge of the canine world”).

<sup>3</sup> See e.g. RICHARD H. FALLON, JR., *IMPLEMENTING THE CONSTITUTION* 93 (2001) (describing the “point” of Holmes’s aphorism as the fact that “we often cannot even characterize an act without understanding what motivated it”). But see Schauer, *supra* note 2 at 197 (understanding Holmes’s claim more modestly

*even* a dog can discern the kick from the trip unless it *matters* what the recipient is likely to take the action to be. The logic of the passage actually works in a way that is different from how it is commonly interpreted: If even a dog knows the difference, then *a fortiori* a person can be expected to know as well. Holmes therefore concedes that the way an action is likely to be perceived matters to the determination of what sort of action it is and how we ought to judge that action.

The competing claim –that intention is determinative– plays a central role in antidiscrimination law, both Constitutional and statutory. There are two important ways in which intention is thought to matter. First, intention may matter because it is the actor’s intention that determines the nature of the action itself – is it kicking or tripping, for example. Second, intention may matter to the determination of the moral or legal permissibility of the action once defined. In antidiscrimination law, is it the thought that counts?

This article will explore the question *whether* intentions ought to be relevant to the assessment of whether a law or policy wrongfully discriminates. In exploring this question, I will use examples both hypothetical and real. When I use case law – as examples to test our moral intuitions – I draw more heavily on constitutional examples. This is because the project began as an attempt at understanding the moral underpinning of the Equal Protection Clause. However, to the extent that the argument focuses on the normative question of the relevance of the actor’s intentions to the moral permissibility of discrimination, its conclusions will also be relevant to discussions about how statutory protections from discrimination ought to be crafted. To reiterate, the subject of this

---

as pointing to the fact that the “distinction between what an agent intends and what actually occurs is an important feature of law generally”).

article is unabashedly moral. It addresses the question whether intentions matter morally. It is my hope that the answer to this question will be taken to be relevant to views about the meaning of the Equal Protection Clause and to discussions of what statutory protections against wrongful discrimination are warranted.

This paper will argue that the actor's intent in enacting a law, adopting a policy or making a decision about an individual is not relevant to the moral assessment of whether the law, policy or decision wrongfully discriminates. While intent and intentions are often prominent features of Equal Protection doctrine and statutory antidiscrimination law, it is not always clear what is meant by "intentions." The first task is therefore to clarify what I mean by "intentions." The actor's intention is what he or she is *aiming at*. To intend is to exercise one's will. It is something over which the actor has control – one does not intend inadvertently, accidentally or unconsciously. One intends *intentionally*. Our intentions are the fruit of our autonomy. Intentions are different than motives as well, though I also believe that the motives of the actor aren't relevant to the moral permissibility of an action either. Though the terms "intention" and "motive" are often used synonymously, "motive" generally refers to a desire-state – that which actually moves the actor to act.<sup>4</sup>

Intentions are not the equivalent of whatever goes on in the head of the actor. Thus, in claiming that the actor's intentions are not relevant to the moral question whether a law, policy or decision wrongfully discriminates, I am not claiming that the mental processes of the actor are irrelevant. Rather I am claiming that what she aims at,

---

<sup>4</sup> See e.g. Steven Sverdlik, *Motives and Rightness*, 106 ETHICS 327, 335(1996) (arguing that "motives belong to the conative, desiring side of the causal story").

or intends to do in acting, is irrelevant. This more circumscribed claim is quite controversial nonetheless.

The role of intention in Equal Protection doctrine and statutory antidiscrimination law has been criticized repeatedly over the years<sup>5</sup> and yet it has endured. I will thus devote a significant portion of the paper to examining what I take to be the two most important arguments for the normative importance of intentions and reply to them in detail. First, however, I begin with the positive case for the irrelevance of the actor's intentions to the wrongfulness of wrongful discrimination. In Part I, I draw an analogy to a debate in the philosophical literature about the Doctrine of Double Effect in order to press the point that the focus on intentions in antidiscrimination law confuses assessment of the wrongness of the action with assessment of the moral blameworthiness of the actor. Intentions do *seem* relevant in many cases. The approach I propose helps to explain this seeming relevance but recasts the relevant inquiry in objective terms. Rather than asking what the actor intends, we ought to ask whether certain features of the law, policy or decision violate norms of fair treatment. The explanation and defense of the objective approach I endorse will be presented in Part I.

Parts II and III will focus on objections to the arguments advanced in Part I. Because Larry Alexander, by himself or with co-author Kevin Cole, has been such a thoughtful defender of the relevance of the actor's intention to the determination of what

---

<sup>5</sup> See e.g. David A. Strauss, *Discriminatory Intent and the Taming of Brown*, 56 U. CHI. L. REV. 935 (1989) (criticizing the discriminatory intent standard in Equal Protection doctrine); Barbara J. Flagg, "Was Blind But Now I See: White Race Consciousness and the Requirement of Discriminatory Intent", 91 MICH. L. REV. 953 (1993) (arguing that the intent standard of Equal Protection doctrine fails to see the way in which seemingly neutral criteria embed a white perspective); Charles R. Lawrence III, *The Id, The Ego and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317, 322 (1987) (arguing that the intent standard of Equal Protection doctrine ignores "unconscious racism"). See also Linda Hamilton Kreiger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161 (1995) (arguing that Title VII's focus on intent is misconceived given current understandings of cognitive bias).

constitutes wrongful discrimination, I will –with his encouragement – use his views as a foil. Part II will explore the argument that an intent requirement is necessary in order to know the real rule at work in a particular policy or law. Like the argument that we need to consult the actor’s intentions to know whether we are dealing with an instance of kicking or an instance of tripping, Alexander believes we need to consult the actor’s intentions in order to know – at least in some cases – whether a law classifies on the basis of a suspect trait or not. In this sense, the intention of the actor defines the nature of the action itself.

Alternatively, one might claim that intention is relevant to the proper evaluation of the law or policy at issue. Where a clearly invidious purpose motivates the enactment of the law or policy, perhaps this invidious purpose (alone or combined with a bad effect) renders a law, policy or decision impermissible. Justice Brennan articulates this view in his 1973 opinion in *U.S. Dept. of Agriculture v. Moreno*<sup>6</sup> when he claims that “if the constitutional conception of ‘equal protection of the laws’ means anything, it must at the very least mean that a *bare* congressional *desire* to harm a politically unpopular group cannot constitute a legitimate governmental interest.”<sup>7</sup> Part III will thus examine whether invidious intent ought to be relevant to our evaluation of whether a law, policy or individual decision is wrongful.

One caveat before beginning. Taking to heart the changes in law review culture as evidenced in the new preference for shorter articles, this article addresses itself to the

---

<sup>6</sup> 413 U.S. 528 (1973) (striking down a provision of a federal food stamp program that limited assistance to households of “related” persons” on the grounds that the statutory limitation was unrelated to the goals of the program (to raise nutritional levels of poor households and to support agriculture) and because there was evidence in the legislative history suggesting that the classification had been inserted to keep hippies and hippie communes from applying for food stamps).

<sup>7</sup> 413 U.S. at 534 (emphasis added).

reader familiar with antidiscrimination law – constitutional as well as statutory. It therefore does not open with a descriptive section whose aim would be to document the ways in which intention continues to play an important role in current antidiscrimination law. The few examples from the case law below are meant as illustrations only, to direct the reader’s attention to the sorts of cases and issues this article will address. They are not offered as *proof* that intent continues to be relevant in our law. What follows in Parts I, II and III are the arguments addressed to the question whether intent ought to be relevant that will, I hope, add to our understanding of what makes discrimination wrong.

#### Part I: The Case Against Intention

Are the actor’s intentions necessary to *define* or to *evaluate* whether a law, policy or decision “discriminates?” I put the term “discrimination” in scare quotes because there is an important ambiguity in the term itself that we must disambiguate before proceeding. The term “discrimination” can be used in a *descriptive* or a *moralized* way. Descriptively, to “discriminate” is merely to draw distinctions among people on the basis of the presence or absence of some trait. For example, the requirement that one must be 16 to drive discriminates between people under 16 and those 16 and over. The requirement that you must pass the bar exam to practice law discriminates between bar passers and non-bar passers, etc. When the term “discrimination” is used in a moralized way, it means to *wrongfully* draw such distinctions. For example, the requirement that group homes for the mentally disabled must seek zoning variances while other group

homes need not wrongfully discriminates between the mentally disabled and others.<sup>8</sup> In order to avoid confusion about which sense of discrimination I am discussing at any particular time I will avoid the terms “discriminate” or “discrimination” altogether. Instead, when I have in mind the descriptive sense of discrimination, I will instead say that the law, policy or decision “classifies on the basis of” X trait – or something similar. When I have in mind the moralized sense of discrimination, I will say that the law, policy or decision “wrongfully discriminates” so that the moralized aspect is explicit.

#### Part I.A: Intention and Identification

With that clarification in place, let’s turn to the first question above: do we need to know the actor’s intentions in order to *identify* whether a law or policy classifies on the basis of some particular trait? Here, it is the descriptive sense of discrimination that is relevant. The idea behind the claim that intentions are determinative here is that classifying itself is an intentional act; that one can only classify intentionally. Is that so?

There is a large body of psychological scholarship extensively summarized by Linda Hamilton Krieger<sup>9</sup> that suggests that it is not. Briefly she claims that “because race, ethnicity, and gender have been made salient by our history and by observable patterns of economic, demographic, and political distribution, people will continue to categorize along these lines”<sup>10</sup> and therefore, “we can expect the resulting categorization-related distortions in social perception and judgment to bias intergroup

---

<sup>8</sup> This example clearly is derived from the case of *Cleburne v. Cleburne Living Center, Inc.*, 473 U.S. 432 (1985) (striking down a municipal ordinance that required special use permits for homes for the mentally retarded that were not required for other group homes).

<sup>9</sup> Krieger, *supra* note 5.

<sup>10</sup> *Id.* at 1239-40.



decisionmaking.”<sup>11</sup> While the extent of cognitive bias may be subject to dispute – and is surely something about which I have no expertise – it is hard to imagine that someone could deny the possibility of the sort of unconscious distortions in perception and judgment that Krieger describes. And, if unconscious bias is possible, what ought we to say about it? If a stereotype about gender, for example, affects the way that an employer makes sense of the data about a particular employee, then gender is – almost by definition – affecting the decision-making process. The employer is classifying on the basis of gender, just not consciously or deliberately. Once we strip the normative evaluation from our understanding of the descriptive inquiry, it is clear that unconscious classification is not only possible but is likely to be prevalent.

Another way to approach the question of whether one can classify unintentionally is to ask a more general question about what sorts of actions can only be done intentionally. After all, the claim that one must consult the actor’s intentions in order to determine if she is classifying on the basis of a particular trait amounts to the claim that classification is the sort of action that can only be done intentionally. First, we should note that the claim that there is a gap between intentions and actions is uncontroversial. The fact that one intends to do a particular action doesn’t guarantee that the intended action will occur. This is true of mundane actions: the fact that I intend to write a persuasive article surely will not ensure that this article is indeed persuasive. And it is true about classification. I may intend to classify on the basis of ability, let’s say, but may not succeed in doing so for any number of reasons.

But what of the opposite direction? Here too, it is commonly the case that the nature of the action doesn’t depend on the intentions of the actor. I can kill someone

---

<sup>11</sup> *Id.* at 1240.

without intending to. I can shoot a basket or hurt your feelings without intending to. However, there are some actions where the actor's intentions seem constitutive. Perhaps one can only murder intentionally. And Holmes' kick distinguishes itself from the trip by virtue of the actor's intentions. Is classification like either of these examples? I think not. Murder is a moralized concept; it is to kill wrongfully where part of what makes it wrongful derives from the culpable mental state of the actor. Kicking is also a misleading example. Sometimes kicking means no more than "to strike out with the foot or feet, as in anger, or in swimming, dancing, etc."<sup>12</sup> This sort of kicking need not be done intentionally; surely the best dancing is done more by instinct. However, when what we mean by "kick" is to strike with the foot intentionally, then this is just part of the way we are using the word – as a sort of short hand for "kick with the foot intentionally." If so, this does not suggest anything deep about kicking, it is just to say that we can use the work kick to mean intentionally striking.

There may well be other sorts of actions that can only be done intentionally, but why think classification is one of them? Rather, there are good reasons to think it is not. Orchestras use a screen to audition musicians in order to protect against the sort of cognitive bias that Krieger describes. If you are the orchestra director and aware of the phenomenon of cognitive bias, and thus on guard against it, why still use a screen? A well-motivated director might choose to use a screen because it is so hard to avoid these errors in perception and judgment. Cognitive bias is hard to root out because it is so easy to classify unintentionally. Not only is classification not one of the unusual sorts of actions that can only be done intentionally, it is the sort of action that is often and easily done unintentionally.

---

<sup>12</sup> Webster's New World College Dictionary, 4<sup>th</sup> Ed.

## Part I.B. Intention and Evaluation

If the actor's intentions do not determine whether a law, policy or decision *classifies* on the basis of a particular trait, perhaps the actor's intentions determine if this classification is wrongful. This claim seems more promising. While intentions may not determine whether an actor classifies on the basis of race or sex, for example, perhaps her intentions matter to the determination of whether this classification constitutes *wrongful discrimination*. It is that claim I will address here. I want to do so however a bit circuitously. The claim that an actor's intentions can affect whether or not an action is permissible or impermissible plays a prominent role in what is known in the philosophical literature as the Doctrine of Double Effect [DDE]. This doctrine – and the claim that underlies it – has generated significant controversy. Because much of the debate focuses on whether the actor's intentions can determine the wrongfulness of the action, a closer look at that debate may be useful here.

According to the Doctrine of Double Effect, actions that will cause harm are sometimes wrong and sometimes not depending on whether the harm is intended or instead merely foreseen. The DDE originates in Catholic Just War doctrine, and thus the example most often used to illustrate it in the secular, philosophical literature is the wartime contrast between the strategic bomber and the terror bomber.<sup>13</sup> The strategic bomber intends to bomb a legitimate military target (a munitions factory, for example) but foresees that in doing so nearby civilians will also be killed. The terror bomber, by contrast, intends to kill civilians. In order to make the cases equivalent except for the difference in the actor's intentions, the example generally supposes that the war is a just

---

<sup>13</sup> See e.g. WARREN QUINN, *MORALITY AND ACTION* (1993) at 177.

war, the same number of civilians will die in each case, and the probability that the civilians will die is also the same in both cases. According to the DDE, the strategic bombing is morally permissible while the terror bombing is not – an intuition that is supposed to be widely shared – *because* the strategic bomber does not intend for the civilians to die, he merely foresees that they will while the terror bomber intends the death of the civilians. As Judith Thomson restates the doctrine (which she calls the Principle of Double Effect [PDE]), “PDE tells us that if the good effect of the act is proportionately good enough, then an agent may morally permissibly perform the act if, while foreseeing the bad effect, he intends only the good effect and does not intend the bad one, either as an end (i.e., for its own sake) or as a means to the good effect.”<sup>14</sup> Because the strategic bomber does not intend for the civilians to die either as an end in itself (because their deaths would demoralize the enemy, say) or as a means to some other end, bombing the munitions factory is permissible. The death of the civilians is truly collateral damage – a regrettable side effect of the legitimate aim. The terror bomber by contrast aims to kill the civilians. Because his intention is to bring about their deaths, his action is impermissible.

A second and more controversial example in which the DDE is used to explain and justify a supposed moral difference between a pair of cases can be found in the area of physician-assisted suicide. Here critics of assisted suicide insist on the importance of the distinction between injecting drugs to hasten death and *merely* refusing life-sustaining treatment. While most of the disagreement in this area focuses on the important and often related distinction between *doing* and *allowing* or in this area specifically between

---

<sup>14</sup> Judith Jarvis Thomson, *Physician-Assisted Suicide: Two Moral Arguments*, ETHICS 109 (April 1999): 497-518 at 510.

*killing* and *allowing to die*, there are some cases to which the *doing/allowing* distinction is inapt and the *intending/foreseeing* distinction more promising. In particular, critics of the moral (and legal) permissibility of assisted suicide often want to draw a distinction between injecting drugs to bring about death and injecting drugs to provide pain relief, where the amount or type of medication required to provide sufficient palliation will predictably bring about death (often called “terminal sedation”). To those for whom these cases are meaningfully morally different, this difference can be accounted for using the DDE. In the case of injecting drugs to bring about death, the actor’s intention is to kill the patient. By contrast, in the terminal sedation cases, the aim of the actor is to relieve the patient’s suffering. The death of the patient, while a foreseeable consequence of injecting the palliative drug, is thus merely a consequence of a legitimate goal.

Justice Rehnquist relied on precisely this account in explaining why the decision in *Vacco v. Quill*<sup>15</sup> sustaining the New York law forbidding assisted suicide would not thereby make terminal sedation illegal. He began by asserting that the legal system in general uses an actor’s intentions as a way to distinguish between actions.<sup>16</sup> In support of this claim, he relied on examples from criminal law in which the actor’s state of mind determines the level of offense<sup>17</sup> and, perhaps because *Vacco* is an Equal Protection case, on that doctrine’s insistence that drawing distinctions among people is only impermissible when a policy is adopted “because of” and not merely “in spite of” their unintended but foreseen consequences.”<sup>18</sup> Drawing explicitly on the DDE, Rehnquist

---

<sup>15</sup> 521 U.S. 793, 802-809 (1997).

<sup>16</sup> *Id.* at 802 (asserting that “[t]he law has long used actors’ intent or purpose to distinguish between two acts that may have the same result”).

<sup>17</sup> *Id.*

<sup>18</sup> *Id.* at 802-3 (citing *Personnel Administrator of Mass. v. Feeney*, 442 U.S. 256, 279 (1979) (upholding a Massachusetts law which accorded veterans an “absolute lifetime” preference in civil service jobs). The law was challenged on the ground that because the class of veterans was overwhelmingly male, the law

argued that terminal sedation is meaningfully different from assisted suicide because the state “may permit palliative care related to that refusal [of medical treatment], which may have the foreseen but unintended ‘double effect’ of hastening the patient’s death.”<sup>19</sup>

Judith Thomson, a critic of the DDE’s insistence that it is the actor’s intentions that matter morally, hypothesizes that the reason that many people continue to believe that intent ought to be relevant (in the cases she examines and in the doctrine of double effect literature more generally) is that they confuse a judgment of the *actor* with a judgment about the *action*. As Thomson puts it, “a failure to take seriously enough the fact – I think it is plainly a fact – that the question whether it is morally permissible for a person to do a thing is just not the same as the question whether the person who does it is thereby shown to be a bad person.”<sup>20</sup> To illustrate this point, consider the following example – drawn from Thomson. Suppose an ill patient is in terrible pain. In addition, suppose there is a drug [D] that will relieve the patient’s suffering. The doctor treating this patient mistakenly believes that in order to provide enough of D to get the patient’s pain under control, D will kill the patient. Actually, D is not very toxic and can be taken in large quantities without any significant ill effects. In addition to being somewhat incompetent, the doctor is also nefarious. The patient is a former colleague of his of whom he was always jealous. As a result, the doctor would like to kill the patient and intends to do so under the cover of providing palliative care (or so he thinks). The doctor

---

constituted unlawful sex discrimination. Justice Stewart, writing for the majority, held that “‘Discriminatory purpose,’” however, implies more than intent as volition or intent as awareness of consequences. It implies that the decision-maker, in this case a state legislature, selected or reaffirmed a particular course of action at least in part ‘because of,’ not merely ‘in spite of,’ its adverse effects upon an identifiable group.” *Id.*

<sup>19</sup> 521 U.S. at 807, note 11.

<sup>20</sup> Thomson, *supra* note 14 at 517. Tim Scanlon endorses a similar view. See T. M. Scanlon, *Intention and Permissibility*, Proceedings of the Aristotelian Soc. Sup. Vol. 74 (2000), 301-317 at 306 (agreeing “with Thomson that this confusion involves a failure to distinguish clearly between features that make an action wrong and description of the flaw that an agent exhibits in performing it”).

is clearly a bad person. His bad intention (to kill the colleague of whom he is jealous) makes our judgment of him an easy case. Less clear is what we should say about the moral permissibility of the action. If intentions matter to determining the permissibility of actions, then we have a strong reason to conclude that the doctor ought not to give the patient the drug. Not only that, we have a strong reason to conclude that another doctor can do the very same act – giving this patient this drug – that the first doctor may not do. These conclusions do not seem right. While the doctor ought not to have the intention he has and is thereby shown to be a bad person, giving the patient the drug is nonetheless the right thing to do. Providing palliation without adverse consequences, which this drug will do, is what the doctor ought to do for his patient.<sup>21</sup>

Thomson's example, while persuasive, may not be exactly on point as the factors that affect whether the physician ought to give the drug are plausibly all external to the mind of the actor: whether the patient has consented, whether the patient is in pain, etc. When we are evaluating whether an actor wrongfully classifies, this is not necessarily the case. While some factors that may affect whether this classification is wrongful are external to the mind of the actor (like the effect of doing so), others are not. Another way to put this point is to note that in classification cases, there is often no single right outcome (analogous to giving the patient the drug). Rather, several policies or decisions are permissible if adopted for permissible reasons. This fact – that it is reasons that are permissible or forbidden – appears to make the actor's intentions crucial.<sup>22</sup>

---

<sup>21</sup> Thomson discusses this same example. See Thomson, *supra* note 14 at 516 (arguing that the fact that the doctor will inject the drug for a bad reason “matters morally, not by way of fixing that it is morally impermissible for her to proceed, rather by showing something morally bad about *her*”).

<sup>22</sup> This is the view of Larry Alexander, which will be discussed in depth in Part II below.

The fact that reasons are central to the determination of whether a decision or policy is permissible need not indicate however that the actor's intentions determine whether an action is wrongful. Here again an analogy to the DDE literature is illuminating. Tim Scanlon, like Judith Thomson, argues against the view that the difference in the actor's intentions explains the difference between the typically paired cases in the DDE literature. In order to explain Scanlon's account, a little background is necessary. For Scanlon, a moral principle contains within it a specification of the reasons that count as exceptions to it. If we follow the principle, we take the reasons that count as exceptions to the general statement as guides to action. If we fail to take these reasons as guides to actions (i.e. act with different intentions) we act wrongly. But, as Scanlon explains, "what makes our action wrong is not the reason on which acted (i.e. our intention) but, rather, *the feature of the action and its circumstances* that the principle identifies as decisive reasons against it."<sup>23</sup> Scanlon provides the following example, which clarifies the rather subtle distinction he is making.

Suppose that I have promised to do something, and that under the circumstances this counts as a decisive reason for doing it. In particular, the fact that I could benefit financially from breaking the promise is not a sufficient reason to fail to keep it. But suppose I break it anyway, in order to get this benefit. In describing what was defective about my action, you might say that I acted wrongly in taking my own advantage as sufficient reason to break my promise. But, at a more fundamental level, what made my action wrong was not the reason I acted on but the reason counting against so acting. The act was wrong because the fact that I promised made it the case, under the circumstances, that I should do the thing in question.<sup>24</sup>

An analogous account can explain the wrongfulness of wrongful discrimination cases as well. Suppose an employer prefers white candidates. He consciously

---

<sup>23</sup> Scanlon, *supra* note 20 at 311 (emphasis added).

<sup>24</sup> *Id.*



and deliberately chooses white candidates over candidates of color. Following Scanlon's account, what makes this wrong is not the fact that the employer *intends* to hire whites. Rather, what makes the action wrong is that skin color ought not to be a factor affecting the hiring decision. It is not the intention that is relevant, rather it is the fact that the race of the job candidate affects a decision when it should not.<sup>25</sup>

Now pair this account with Krieger account of cognitive bias. In the above section, we established that people classify on the basis of race, sex or other traits without intending to. In other words, race or sex based traits influence decision-making without cognitive awareness. And, if what makes an action wrong is the fact that race is playing a role it ought not to play, then what makes such an action wrong is not the actor's intentions but the objective fact that racial categorization played a role in the decision-making that it ought not to have played.

It is important to pause over this point because it is this sort of case that is at the root of the view that intention must be relevant to the evaluation of claims of wrongful discrimination. Most people agree that where an action is morally required, tossing the life preserver to the drowning person, for example, the action

---

<sup>25</sup> It might be tempting to draw the analogy even more closely to Scanlon's and argue that what makes the hiring decision wrong is that a particular job applicant objectively ought to be hired. This is a mistake however. In the context of jobs or places at schools, it is not the case that particular candidates are entitled to those spots. The employer or school can choose whichever applicants it wishes. Rather, morally as well as legally, the employer or school is merely proscribed from basing its decision on certain characteristics in certain contexts or in certain ways. While the issue of affirmative action shows the complexity of spelling out when or in what context race may be a legitimate criterion in job selection or school admission, it is probably safe to say that an employer or school may not disfavor minority applicants on the basis of race. In this article, I do not provide the basis for this moral judgment – which would be necessary for a complete account of wrongful discrimination – rather I hope to build on what I take to be a widely shared moral view. Taking as the starting point then that the use of race to disadvantage a minority applicant is wrong, an employer who does so acts wrongly *not because* she intends to disadvantage the minority applicant but because she uses race in her decision-making in a way she should not.

is right whether one does it in order to save the person or in the hope of a reward. Discrimination seems different however. This is because wrongful discrimination often occurs in contexts where no one particular action is required. In the employment context, often no one particular person is clearly the best qualified, for example. In such cases, it often seems that what is forbidden therefore is acting with a particular intention (hire only whites) or from a particular motive (racism). Scanlon's insight – although it does not deal with a case of multiple permissible actions – exposes the confusion behind this intuition. Although multiple actions are permissible, some actions – properly described – are forbidden.

Steven Sverdlik, arguing in a similar vein as Alexander for the moral relevance of the actor's motives,<sup>26</sup> thinks that discrimination presents one of the kinds of cases in which motivation has moral significance. He uses the example of a seller who refuses to sell his house to a prospective buyer because the buyer is black. Sverdlik believes that because the seller could refuse to sell to the buyer for other reasons (lack of credit worthiness, for example) or to take his house off the market altogether (because he has second thoughts and decides to keep the house) and thus is not obligated to sell his house to this would-be buyer, it must be the case that the actor's motives make the moral difference – turning a morally permissible action (refusing to sell for the two later reasons) into a morally wrong action (refusing to sell because the buyer is black). But is it really the actor's motives that matter?

---

<sup>26</sup> Sverdlik, *supra* note 4 at 341-9.

Where multiple actions are permissible – selling or refusing to sell to this buyer, for example – reasons do matter but let's look more closely at how they matter. What is morally impermissible is for the race of the buyer to affect the seller's decision. Thus the seller may decide to sell or not to sell for a myriad of permissible reasons but he may not refuse to sell to the buyer because of the buyer's race. If the buyer's race affects the seller's decision, a wrong has occurred. If motives are a species of desire and intentions are a product of the will, then neither motives nor intentions matter in that it doesn't matter whether the actor desires or is motivated by race nor whether he intends or wills that race play a role in his decision-making process. What matters is whether indeed the race of the buyer was a factor in his decision, whether he wanted it to be or intended it to be or not. If race played such a role, the action is wrongful.

Intending is an act of will. The claim that one must intend to wrongfully discriminate in order to do so maintains that to classify wrongfully one must be intending to do so. But if one can *classify* unintentionally, why think that *wrongful* classification requires an intent to classify? This view confuses a judgment of the actor with a judgment of the action, in the way that Thomson describes. The wrongfulness of the action isn't a statement about the virtue of the actor; it is a statement about the action. While this article has not articulated a positive account of what – rather than bad intentions – renders a classification morally problematic, very roughly it seems plausible to say that wrongful discrimination fails to treat people with equal concern or fails to provide fair treatment, or something of this nature. Here what is emphasized is the treatment people receive. In other words, the argument against the importance of the

actor's intentions breaks the link between the actor's autonomy and the wrongfulness of the action. Sometimes we treat people unfairly even though we do not mean to do so. Often, it's not the thought that counts.

## Part II: Is Intention Necessary to *Identify* the Classifying Trait?

Larry Alexander offers an argument against the view I present above. He believes that the actor's intention is necessary in order to know what sort of action one is dealing with in the first place. His complex and challenging article *Rules, Rights, Options and Time*<sup>27</sup>, in which he lays out his argument for this view, thus presents a useful foil. In what follows below, I will briefly describe Alexander's argument and then reply to it.

Alexander argues (adopting Matt Adler's conception of constitutional rights as rights against rules<sup>28</sup>) that there are many constitutionally permissible rules that have the effect of excluding people. For example, the rule "the pool is now closed" excludes those who want to swim now.<sup>29</sup> (I use and build on this example because it is one that Alexander discusses.) In addition, he argues that it is also permissible for a state actor to change his or her mind about which of these constitutionally permissible rules to adopt. For example, "pool open" and "pool closed" are both constitutionally permissible rules. However, if different rules impact different groups of people adversely, then switching

---

<sup>27</sup> Larry Alexander, *Rules, Rights, Options, and Time*, LEGAL THEORY 6 (2000), 391-404.

<sup>28</sup> See Matthew D. Adler, *Rights Against Rules: The Moral Structure of American Constitutional Law*, 97 U. MICH. L. REV. 1 (1998).

<sup>29</sup> I don't intend to myself adopt a view about whether Adler's conception of constitutional rights as rights against rules is correct, nor does Alexander. Nonetheless Alexander's point that many constitutionally permissible rules exclude people is surely correct – however one conceives of constitutional rights. This fact – which one might call the *discrimination conundrum* – was first articulated by Tussman and tenBroek in their classic article *The Equal Protection of the Laws*. See Joseph Tussman & Jacobus tenBroek, *The Equal Protection of the Laws*, 37 CALIF. L. REV. 341 (1949). In order to accomplish a myriad of both important and mundane tasks, laws and policies must draw distinctions among people (must "discriminate" – to use that term without its normal pejorative connotations).

between constitutionally permissible rules in a *calculated* manner may produce troubling results.

For example, suppose the person in charge of determining when the town pool is open or closed opens the pool when he sees a group of white patrons approaching and closes the pool when he sees a group of black patrons approaching. Alexander argues that here the *real* rule is that whites may swim and blacks may not. In his view, “[i]t is the operative rule for constitutional assessment.”<sup>30</sup> And, and here’s the key claim, the only way to know that this is the real rule is by looking to the actor’s intention: “the purpose essentially defines the real rule that must be constitutionally evaluated.”<sup>31</sup> While the term “purpose” itself may be ambiguous – purpose may refer to the subjective motivation of the actor or purpose may refer to the best understanding of the policy’s aim, viewed objectively – Alexander is clear that it is the internal motivation of the actor that for him is key. Alexander’s analysis requires that a court make a determination about the content of the actor’s mental state. He says repeatedly that it is motivation that matters<sup>32</sup> and emphasizes the reasons *for which* a policy was adopted.<sup>33</sup> In other words, it is the actual motivation of the actor, the reasons he adopted a policy, that count rather than the reasons that support a policy being adopted.

This is an important argument. If indeed we need motivation analysis to distinguish this case from instances of benign switching between constitutionally permissible rules, then Alexander is right and it is the thought that counts. But we do not.

---

<sup>30</sup> Alexander *supra* note 27 at 400.

<sup>31</sup> *Id.*

<sup>32</sup> *Id.* (arguing that “[g]iven constitutional optionality and the permissibility of switching, the courts must treat legislative motivation as material).

<sup>33</sup> *Id.* at 400-01 (claiming that “it is material whether they [laws] were enacted for reasons that were also constitutionally optional as opposed to constitutionally forbidden”).

Below I present two arguments against Alexander's contention that motivation analysis is necessary to identify cases like the one he describes above.

Before doing so, however, it is important to clarify Alexander's position. Alexander claims that the "purpose essentially defines the real rule."<sup>34</sup> In other words, we need to look to purpose to know what the rule is because intentions are constitutive of the action. This view should not be confused with a more modest epistemological claim. Someone might believe that whether or not a person classifies on the basis of race, sex or some other trait is determined independently of that person's intentions and yet believe that consulting the actor's intentions is a fairly good way to find out about the real nature of the action. If people generally do what they intend to do, this may well be reasonable. Nor should Alexander's view be confused with a doctrinal claim: that an intent-based test reaches the right result (where rightness is defined independently) more often than any plausible alternative.<sup>35</sup> Alexander's claim about the relation between intentions and classification is far more basic and constitutive.

#### Part I.A. Intention Is Neither Necessary Nor Sufficient

Alexander argues that we need intention analysis to *determine* what he terms the "metarule" (here, that whites may swim but blacks may not). The first point to note about Alexander's argument is that it is circular. He assumes that the real rule *is* what the actor intends. That is why he tells us that the lifeguard has been instructed to display the pool open sign when whites arrive and the pool closed sign when blacks arrive. But if the

---

<sup>34</sup> *Id.* at 400.

<sup>35</sup> See e.g. the distinction between constitutional "operative propositions" and "decision rules" proposed by Mitchell Berman, Mitchell N. Berman, *Constitutional Decision Rules*, 90 VA. L. REV. 1 (2004) (distinguishing "operative propositions" which elucidate constitutional meaning from "decision rules" which allow courts to apply that meaning).

example is offered as an *argument* for why intentions are relevant, we must *start* from a place that does not yet make any assumptions about what the real rule is.<sup>36</sup> A fairer way to present the issue would be to consider the following three scenarios and then ask oneself whether intentions matter as Alexander supposes.

*Scenario 1:* A lifeguard on duty displays the sign “pool open” when whites arrive to swim and the sign “pool closed” when blacks arrive to swim.

*Scenario 2:* Same as above, plus the lifeguard is switching the rule in order to keep blacks out.

*Scenario 3:* Same as 1, plus the lifeguard is not switching the rule in order to keep blacks out.

If no other plausible explanation is offered (as Alexander’s example assumes), the rule in all three cases may be the same. There are two points here. First, we don’t need to know the actor’s intentions to define the meta-rule. Sometimes it is determined from what takes place. Second, the *real rule* can be different from what the actor intends. In Scenario 3, notwithstanding the fact that the actor may not be racially motivated, if he consistently opens the pool for whites and closes it for blacks, then the real rule just is that the pool is open for whites but not for blacks. The more persistent the pattern, the clearer this is.

In making this claim, am I thereby endorsing disparate impact as the test for what the rule is – and thus for what counts as wrongful discrimination? Yes and no. Yes, in that on the analysis I offer, it is the objective features of the action, law or policy that

---

<sup>36</sup> To be clear, let me explain the relationship between Alexander’s term “real rule” and the vocabulary I have been using: classifying on the basis of X trait and wrongfully discriminating. When Alexander asks what the real rule is, he is asking about X. In other words, he is interested in the descriptive sense of discrimination, about classifying on the basis of X trait but he is raising a question about what is the X on the basis of which the classification is made. To identify the “real rule” is to find the X on the basis of which the rule classifies.

matter and disparate impact is one such objective feature. But no, in that this approach does not commit me to the view that disparate impact is the *only* or indeed the *central* objective feature that matters. Indeed, even under an intent-based approach disparate impact is relevant as evidence of subjective intent. In my view, disparate impact is evidence of objective meaning. The Supreme Court's disparate impact cases suggest that impact alone cannot be sufficient to establish intent – though it is unclear, at least theoretically, why not. On my view, disparate impact alone could establish the objective meaning of the law but in the kinds of cases that have come before the Court – like *Washington v. Davis*,<sup>37</sup> -- there is other important objective evidence that cuts in the other direction. For example, the fact that there are good reasons to test police officers for the skills the test in question evaluated is relevant to the objective meaning of the law. Note that I am not suggesting one ask whether these reasons actually motivated the police in adopting the test.

So far, I have argued that objective features of the law or policy in question permit us to determine what the real rule is and thus that intention analysis is not necessary. The example we have focused on (pool open when whites arrive and closed when blacks arrived) presented a fairly unambiguous case where the objective indicia pointed clearly to the real rule: whites may swim but blacks may not. Alexander might respond to this critique by pressing a more ambiguous case. Consider the following example, drawn from another excellent and thought-provoking article by Larry

---

<sup>37</sup> 426 U.S. 229 (1976) (rejecting a challenge to the use of a written test in hiring decisions for the D.C. Police Department on the grounds that the fact that the test had a disparate impact on African-American job seekers was not sufficient to shift the burden to the Police Force to show that the test was a good predictor of job performance).



Alexander (here with Kevin Cole).<sup>38</sup> A hypothetical state supported law school in the plain states adopts a policy of favoring regional applicants. It does so however because this policy will decrease minority enrollment, which the school would like to do because it turns out that the LSAT overpredicts the performance of minority applicants (these are the facts assumed by Alexander and Cole in their examples). By decreasing the number of minority students admitted, the law school will thereby enhance the performance of its students.<sup>39</sup>

Two questions are raised by this hypothetical. First, is Alexander right that the *real rule* here is “limit minority applicants” (as an intent analysis might suggest) rather than “prefer local applicants”? If we don’t *assume* that the intention defines the real rule, it is not at all clear that the real rule is “limit minorities.” It is important to distinguish the rule from the motive for adopting the rule. A more natural way to describe this case would be to say that the rule is “prefer local applicants” but that the motive for adopting the rule is to limit minority applicants (in order to boost overall student performance). Where the objective features of the policy do not manifest the internal subjective motivation, to say that the real rule is defined by the subjective motivation is to merely assert his view and at the cost of some fairly counter-intuitive formulations.

The second question raised by this example is more serious. If the rule “prefer local applicants” is adopted in order to limit the enrollment of minority students, does this reason itself provide grounds for holding the policy impermissible? This issue – whether problematic subjective intentions are relevant to the evaluation of a law, policy or decision will be considered in Part III below.

---

<sup>38</sup> Larry Alexander and Kevin Cole, *Discrimination by Proxy*, 14 CONST. COMMENT. 453 (1997).

<sup>39</sup> This is the School C that Alexander and Cole describe, *see* Alexander and Cole, *supra* note 38 at 454.

## Part II.B. Facial Discrimination, Disparate Impact and Subjective Intent

Subjective intention matters for Alexander in order to determine the real rule. By this he means that intention is necessary to determine whether a law or policy classifies on a suspect basis or not. Alexander is not interested in the motive for classifying (in the way that affirmative action is often described as benign discrimination). Rather he is only interested in using subjective intention to determine whether the real rule is one that classifies on a suspect basis or not (for whatever reason). As a result, for Alexander there is a fairly firm line between disparate treatment and disparate impact cases.<sup>40</sup> Disparate treatment cases are not Alexander's focus, presumably because in these laws or policies, the fact that the law classifies is manifest and thus inquiry into subjective intent unnecessary. But can Alexander distinguish between disparate impact and disparate treatment cases in this hard and fast way and cabin the use of subjective intent analysis to disparate impact cases?

To examine this question, let's look a bit more closely at how Alexander detects the "real rule" at work in disparate impact cases. In discussing policies adopted at California universities to promote diversity in the wake of proposition 209 (which prohibits race based affirmative action), Alexander says the following:

Administrators who select facially nonracial admissions criteria 'in order to promote racial diversity' are not avoiding the law but are violating it. They are choosing close proxies for race for racial reasons, and we must assume that if the reasons remain constant over time, the proxies will change if necessary. ... The real rule is 'admit so as to promote racial diversity,' which is forbidden.<sup>41</sup>

---

<sup>40</sup> Alexander's focus is on disparate impact cases largely as he sees these as the sites of the sorts of conceptual puzzles he finds both troubling and illuminating.

<sup>41</sup> Alexander, *supra* note 27 at 33.

In order to determine the real rule, Alexander looks to the subjective intent. If a university wants to admit some percentage of minority applicants but is forbidden from explicitly classifying on the basis of race, the school might instead adopt a facially neutral policy like “admit the top 10% of all high school graduates.” Alexander claims that this policy *is* a racial classification itself. This is the odd and strongly counter-intuitive implication of his view that I highlighted in the previous section. On what *basis* does he determine that this is the real rule, notwithstanding the objective evidence provided by the policy itself? He determines that the real rule is “admit so as to promote racial diversity” because over time the proxies used to achieve this goal will change. In other words, because the top 10% rule is likely to be retained only so long as it serves the diversity goal, we ought to conclude that the real rule is “admit to promote racial diversity” rather than “admit top 10%.”

However, this same analysis can be used to show that rational race or sex discrimination isn't actually race or sex discrimination at all. For example, suppose a state employer refused to hire women because in general women decline to work the long hours that are required by the jobs in question. Here sex is being used as a proxy for willingness to work long hours.<sup>42</sup> Is this policy an instance of discrimination on the basis of sex? Alexander's analysis would seem to suggest that the answer is no. Like in the example explored above, one cannot simply look at the face of the rule or policy, rather one must discover the “real rule.” To do so, according to Alexander one looks at

---

<sup>42</sup> Interestingly, this is the first reason offered by Harvard President Lawrence Summers to explain the under-representation of women as tenured professors in the sciences as well as in other demanding jobs in his remarks at the National Bureau of Economic Research [NBER] Conference on Diversifying the Science & Engineering Workforce that created so much controversy due to the second reason he offered (innate differences in abilities between men and women). *See* [www.president.harvard.edu/speeches/2005/nber.html](http://www.president.harvard.edu/speeches/2005/nber.html).

subjective intent. In particular, one looks at whether “the proxy will change as necessary” to accommodate the underlying purpose. In the case of rational sex discrimination (or rational race discrimination), the underlying purpose is to hire the most cost effective worker. Alexander’s analysis would therefore imply that the real rule is “don’t hire people unwilling to work long hours” rather than “don’t hire any women.” As such, this is simply not a case of sex discrimination at all.

Alexander’s analysis fuses the normative question of whether the discrimination is permissible with the interpretive question of what sort of classification is being used (i.e. what the “real rule” is – to use his terms). To argue that an employer’s refusal to hire women because in general women are unwilling to work long hours is not gender discrimination seems just plain wrong. If the rationality of the classification matters – which it does to some commentators and not to others – surely it relates to the moral or legal permissibility of the classification not to whether this is a gender classification at all.

The laws and policies at issue in disparate treatment cases are often described as “facially” discriminatory. This term is illuminating. The claim that a law or policy is facially discriminatory is to emphasize that objective features of the law (its text) are what matter in determining what the *real* law is. The objective evidence tells us both what the real rule is and provides the material on the basis of which to evaluate its permissibility. Alexander is right to point out that the facial evidence (meaning what the law or policy directs on its face) is not always dispositive. When we claim that the law is facially discriminatory, we are really saying that the facial evidence *outweighs* other evidence. But what other evidence would it make sense for it to outweigh? If evidence

of subjective mental state could trump the facial evidence, we are left with the puzzle of why it doesn't trump it in cases of rational race or sex discrimination. If however, facial evidence is one type of objective evidence that together illuminate the "real rule," then when we conclude that a law discriminates based on the facial evidence, we are saying that no other objective evidence of contradicts or overcomes the important facial evidence of what the law really is.

To summarize the points discussed thus far: first, intention is not necessary to identify the *real* rule at work in a law. Objective evidence will suffice to identify troubling cases. Neither is it sufficient. The pattern of rule switching that is motivated by an intent to keep Blacks out is not necessarily different from one that is not. To insist that it is different is not really to make a claim about the need to consult intentions to *identify* the rule a work but instead to make a claim about the how one ought to *evaluate* a rule that is motivated by bad reasons – a claim that will be explored and evaluated in Part III below. Second, the claim that one needs to look at subjective intent in order to discover the "real rule" leads to the counter-intuitive conclusion that facially discriminatory laws that discriminate rationally (and perhaps others) do not really classify on the basis of sex or race. On this view, the real rule is defined by the proxy's target (hire the most productive employee, for example). This account seems wrongheaded. Either subjective intent matters in both disparate treatment and disparate impact cases to determine what the real rule is, which leads to unsupportable conclusions or objective evidence determines the real rule in both types of case.

Part II.C. Private Bias

One kind of case continues to be troubling to the account I have provided – the sort of case in which the actor uses race or sex (or something else troubling) as a criterion in a way that eludes detection. Imagine that an employer follows the policy of hiring the candidate she judges to be the best qualified but in cases of ties, hires the white job seeker. Because the employer only uses race to break ties, the manifestation of this policy may not suggest that race has played a role. Another employer for whom race was not a factor could easily have chosen to hire the same applicants as the first employer, but for other reasons. It seems that we should say about this example that the “real rule” is that race played a role in the selection process. In order to do so, however, must we consult the actor’s intentions, as Alexander supposes?

In order to answer this challenge, consider two different employers, A and B. Employer A decides that where two job seekers have equivalent relevant credentials, she will favor the white job seeker. In adopting this policy, she makes a deliberate choice that race will play a role in her employment decisions. B, by contrast, chooses the candidate she thinks is best qualified. However, in a way of which she is not consciously aware, she consistently devalues the accomplishments of non-white candidates.<sup>43</sup> The bias is small (which partly explains why it continues, unrecognized by B) so that it leads to the same employees being selected over time as employer A chooses. In these cases, I think it is fair to say that A intentionally uses race as a factor in decision making while B does not. B is not even aware of the fact that race is relevant to his decision making process and he surely has not adopted the preference for non-whites as a reason for action. If, following Alexander, we say that the “real rule” in the case of employer A is

---

<sup>43</sup> For a complex and detailed account of how cognitive bias affects both perception and judgment, *see* Krieger, *supra* note 5.

“choose the best candidate, but in case of ties choose the non-white” *because* this is what employer A intends, then what must we say about employer B? It seems we must say that the “real rule” used by B is “choose the best candidate” – as that is what she intends. Yet, that doesn’t seem right.

If one adopts the perspective of the people affected by each of these policies, it seems clear that the real rule in both these cases uses race as a factor that counts against the applicant. While we might blame the employer in case A and not in case B as the actor’s intentions may be relevant to a judgment of the actor (about this issue I take no stand), that doesn’t mean that the actor’s intentions determine what *action* the actor has indeed taken. Both A and B use race in decision-making. In order to determine what the real rule is we need to know whether race was a factor, not whether the employer intended race to be a factor. Race was a factor for B as it was for A, notwithstanding the lack of B’s intention to make it so.

This discussion helps illuminate what we might mean by the “real rule” – to use Alexander’s locution. An approach that privileges intentions allows the actor’s conscious choices to be determinative of the real rule. However, because the social meaning of an actor’s action can be different from what she intends (discussed earlier) and because cognitive bias can affect perception and judgment, what an actor intends often doesn’t track what the actor does in fact. Therefore, the real rule refers to the factors that actually affect the choice of job candidate – whether the actor intends them to do so or not.

Now of course there remains an important practical problem. How should a court determine what the real rule is? How is anyone to know that race was a criterion in the decision-making for both A and B since the criterion is private, even obscure to the

person using it? The first thing to note about this issue is that it presents a question at the level of doctrinal implementation not about what factors *really* matter to a determination of what the real rule is. As the aim of this paper has been to argue against the claim that intentions *really* matter, establishing this much is sufficient to the philosophical question addressed here.

That said, let me offer some brief thoughts about the question of doctrinal implementation. Someone might argue that a doctrinal approach that focuses on intentions is the best approximation of the factors that really determine what the real rule is – after all in many cases what that actor intends and what she does will be the same. While this approach may initially seem plausible, I think it is far less so that it initially seems. First, intentions themselves are non-public and thus difficult for a challenger to establish. Therefore, this approach doesn't even have the legal virtue of administerability. Second, there are good reasons to think that the intended act and the actual act will often come apart – either because cognitive bias is pervasive, as Linda Hamilton Krieger persuasively argues,<sup>44</sup> or because the social context in which people act has a powerful impact on what that action actually is. Let me end with an example that will illustrate this second point. If the governor of a southern state were to fly the confederate flag over that state's statehouse, the fact that he does so for good or at least morally benign reasons is not enough to determine what he expresses in so doing. The history of slavery, lynching and repression in the South – both before and after the Civil

---

<sup>44</sup> *Id.*



War – imbues that symbol with a meaning that this actor’s intentions cannot fully control. The act has a meaning distinct from the speaker’s intended meaning.<sup>45</sup>

### Part III: Is Invidious Intent Relevant to the *Evaluation* of Classifications?

In this section, we deal with the second sense of *intention* identified at the start of the article. Recall, sometimes courts and commentators are interested in the question: on the basis of what trait does the law or policy classify. The relevant intention thus relates to that question: does the actor intend to classify on the basis of sex or veteran status, on the basis of residency or race. Intention is relevant in this sense in order to *define* what the law or policy to be evaluated actually *is* – thus Alexander’s talk of the “real rule.”

Alternatively, intention may be relevant to *evaluate* a law, policy or decision – once it is defined. The actor’s aim in classifying as she does is the relevant intention here. But still there is an ambiguity. Often a classification is used as proxy for another trait, which I will call the “target.” We could say that the actor’s intention in classifying on the basis of X trait is to identify people with Y trait (the target). The actor’s intention then is to select people with the target, Y, trait. In addition, there is a reason that the actor aims at Y. That reason could also be described as the aim of the actor. He intends to classify on the basis of X trait, in order to pick out people with Y trait, for Z reason.<sup>46</sup>

For example, if a state law school adopts a regional preference in order to reduce the

---

<sup>45</sup> This position is analogous to the view that a sentence or utterance has a meaning independent of the speaker’s intended meaning. For an accessible overview of the distinction between speaker’s meaning and sentence meaning, see Matthew D. Adler, *Expressive Theories of Law: A Skeptical Overview*, 148 U. PENN. L. REV. 1363, 1374 (2000) (explaining that the “speaker’s meaning of a statement is (roughly) what the speaker intended to communicate by her utterance of it; its sentence meaning is (roughly) what the statement is conventionally uttered to communicate).

<sup>46</sup> I explain the terminology “proxy,” “target” and “motive” in more depth in a previous article in which I describe two forms of discriminatory laws – one’s that work by proxy and one’s that do not. See Deborah Hellman, *Two Types of Discrimination: The Familiar and the Forgotten*, 86 CAL. L. REV. 315 (1998).

number of black students admitted, in order to raise student performance, the school uses *the proxy* “local” to aim at *the target* “non-black” because it wants to raise student performance (the purpose or goal).<sup>47</sup> If intention matters, is this because the intended target matters<sup>48</sup> or because the purpose or goal matters? Below, I will consider each in turn.

### Part III.A The Target

In order to isolate whether aiming at a problematic target is relevant to whether a law or policy is permissible, let’s begin with a case in which the law uses a benign proxy to aim at a problematic target for a good (or at least permissible) reason. Alexander’s regional law school provides an apt example. In that case, as you will recall, the school authorities decide to adopt a regional preference in order to decrease the number of black students admitted. This is the problematic target; local student is used as a proxy for the target non-Black. The reason that motivates this decision however is benign. The school has found (through the experience of a law school in a nearby state) that the LSAT over-predicts the performance of black students as compared to white students. Though these facts present an unlikely scenario, it is a useful hypothetical to test our intuitions about whether and how the use of a suspect target matters.

---

<sup>47</sup> The purpose or goal may be aimed at as a means toward another end as well. *See* Mitchell N. Berman, *Coercion Without Baselines*, 90 GEO. L. J. 1, 24 (2001) (describing this relationship as “nested purposes” and endorsing the view that a bad purpose “wherever a putatively bad purpose may lie within a chain of nested purposes, the purposes animating state action, like its effects, constitute familiar respects in which the Constitution can be violated”).

<sup>48</sup> When we focus on the target, this inquiry seems very close to the inquiry discussed in Part I. In this Part, we are analyzing whether it matters that an actor uses a benign trait to target a suspect classification – like race or sex. That is, we are assuming that is what the actor is aiming to do and asking whether that intention ought to have legal significance. This inquiry is not the same as the inquiry into whether one must consult intention in order to *detect* the case in which that actor uses a benign trait in order to target a suspect class.

If we are trying to assess whether the intention to exclude blacks matters to the determination of whether the policy is wrongful, we should first assume that the policy doesn't have its desired effects. Suppose, it is adopted for the reasons specified, but the demographics of the two states are sufficiently different so that this regional policy does not in fact have a negative impact on blacks. Here to hold that the proxy's target renders the law impermissible would seem to fetishize the purity of legislative motive. If we are not in the business of judging the virtue of legislators or other policy makers, this ineffectual attempt to exclude blacks seems irrelevant. More interesting and troubling is a case in which the proxy does reach the desired target – fewer blacks are admitted. In such a case, does the intent to raise student performance via reducing the number of blacks matter to assessing the permissibility of this policy?

Alexander and Cole believe that intent is crucial in such a case because without looking at intent we are unable to distinguish this policy from the benign adoption of a local preference. In other words, Alexander and Cole believe that an intent standard is required to support what they term the “anti-discrimination principle” by which they mean the rule that some racial and sex discrimination is forbidden.<sup>49</sup> If one can *get around* the prohibition on racial and sex discrimination by substituting a close proxy for race or sex, then the prohibition is useless, or so they argue. Thus, Alexander and Cole thus argue that the intent principle “is necessary to prevent easy circumvention of the Anti-Discrimination Principle.”<sup>50</sup>

---

<sup>49</sup> Alexander and Cole, *supra* note 38 at 453 (defining the anti-discrimination principle in the following way: “Government cannot use racial classifications, even as the most cost-effective proxies for other traits, unless using them as the most cost-effective proxies is necessary to further a compelling interest”).

<sup>50</sup> *Id.* at 455.

But is this right? The only way to know whether the school is classifying on the basis of residency in order to restrict the number of blacks (in order to raise student performance) *is* to consult intentions. But this merely states the obvious. The only way to know the school's intentions is to look at the school's intentions. As an argument for why these intentions are relevant, it is a non-starter.

In order to begin to think about whether this intention ought to be relevant to the permissibility of this policy, consider the following variation on Alexander's example. Suppose that the school adopts the preference for local applicants in order to enhance student performance but without recognizing why this works. After the fact however, studies demonstrate that the reason that the local preference enhances student performance is because it lowers the number of minority candidates admitted which in turn enhances student performance. Now what? Should the school repeal the policy? If you are the school official responsible for deciding what to do, how should you think about this question? If Alexander and Cole are correct that it is the intent to use locality as a proxy for race that is problematic, then it would seem that the administrator need not worry in this case. The policy of her school was adopted without this wrong-making intention and so is not wrongful. Remember, I am not considering the issue of what sorts of actions might make one vulnerable to suit in the real world or what prophylactic or defensive actions might well be prudent. Rather, the aim of this article is to consider when drawing distinctions among people is or is not wrongful.

Instead, and taking a page from Scanlon, the school administrator should think about the question in this way: *Is enhancing student performance an adequate reason to prefer local applicants when this policy works by lowering the number of minority*

*candidates admitted?* Now of course, a school administrator who doesn't know that the policy works this way will only be able to ask herself the question: *Is enhancing student performance an adequate reason to prefer local applicants?* Her lack of knowledge may well be relevant to our assessment of praise or blame of her, but it is not relevant to our assessment of the policy itself. If the policy works to enhance performance by limiting minority enrollment, this is a feature of the policy that must figure in our assessment of its permissibility whether it is known or intended by the actor or not. Thus, when this fact comes to light, the administrator judging the permissibility of the policy ought not consult her (or the school's) intentions to determine whether she ought to repeal the policy. Rather she ought to ask the first question posed above: *Is enhancing student performance an adequate reason to prefer local applicants when this policy works by lowering the number of minority candidates admitted?*

But isn't it worse to do reduce minority enrollment *deliberately*? That depends what we mean by "worse." If the answer to the question above is that it *is* wrong to enhance student performance via a local preference that lowers minority enrollment, then to try to evade the law reveals something vicious in the character of the actor. She acts in a way that shows she thinks she is above the law. However, if it is wrong to enhance student performance via a local preference that lowers minority enrollment, then it is wrong to do it unintentionally as well as intentionally – and neither *action* is worse than the other. Similarly, if it is permissible to do so, then it is permissible to do so intentionally.

Alexander and Cole fuse the questions of how we ought to judge the actor and the action by using the term "circumvention." It suggests the crafty evasion of the law that

reveals something distasteful in the actor. If we are careful not to let this distaste for the actor bleed into our judgment of the action, we can address the question of circumvention more carefully. Whether or not the use of locality as a proxy for race as a proxy for student performance is indeed a circumvention of the prohibition on rational race discrimination (that is, on the use of race as a proxy for student performance directly) depends on what it is exactly that is forbidden by a prohibition on rational race discrimination.

At least two possibilities come to mind – each of which would have different implications for the question whether the use of locality as a proxy for race as a proxy for performance ought similarly to be forbidden. Before I discuss them, I want to pause to emphasize that although this example and discussion focuses on rational race discrimination, the same observations and inquiries would apply to rational sex, disability or other forms of rational discrimination. For ease of exposition, I will focus on rational race discrimination. The prohibition on rational race discrimination forbids generalizing over race even when the generalization is supported by available data. Thus, though race may be as good a proxy for a desired outcome (school or job performance, for example) as other indicia like grades or experience, we forbid the use of race as a proxy in most instances. In doing so, however, what exactly do we forbid. At least two possibilities come to mind. Perhaps we forbid a) the use of race as a proxy to disadvantage or b) the explicit or near explicit use of race as a proxy to disadvantage?

The choice of between these two conceptions of what rational race discriminations forbids, (a) or (b), will depend on the reason why one believes rational race discrimination itself ought to be forbidden. If the problem with rational race

discrimination is that it tends to support and entrench the racial hierarchy of our society – an anti-caste understanding of antidiscrimination norms<sup>51</sup> – then the prohibition on rational race discrimination forbids (a). If instead the problem with rational race discrimination is that generalizing over race tends to reify racial categories or expresses denigration of racial minorities,<sup>52</sup> then the prohibition on rational race discrimination may forbid something more narrow, something like (b). These are not meant to be exhaustive formulations or explanations, just exemplary.

If the prohibition on rational race discrimination forbids (a), then the use of locality as a proxy for race as a proxy for performance ought also to be forbidden. Intending to use locality as a proxy for race in this way would indeed be an attempt to circumvent the prohibition on rational race discrimination by carrying it out in a manner that is hard to detect. But, it is not the intent to circumvent that makes it wrong. On this understanding of what makes rational race discrimination wrong, the use of locality as a proxy for race is wrong, whether it is intended or unintended.

On the other hand, if the prohibition on rational race discrimination forbids (b), then the use of locality as a proxy for race as a proxy for performance is not wrongful. A policy that privileges local applicants as a way to enhance student performance does not express denigration of racial minorities or reify racial categories because race is not used

---

<sup>51</sup> See e.g. Owen M. Fiss, *Groups and the Equal Protection Clause*, 5 Phil. & Pub. Aff. 107 (1976); Cass R. Sunstein, *The Anti-Caste Principle*, 92 Mich. L. Rev. 2410 (1994).

<sup>52</sup> Justice Thomas makes an argument along these lines in his opinion, concurring in part and dissenting in part in *Grutter*:

The Constitution abhors classifications based on race ... because every time the government places citizens on racial registers and makes race relevant to the provision of burdens or benefits, it demeans us all. "Purchased at the price of immeasurable human suffering, the equal protection principle reflects our Nation's understanding that such classifications ultimately have a destructive impact on the individual and our society." 539 U.S. at 353-4 (quoting *Adarand Constructors, Inc. v. Peña*, 515 U.S. 200, 240 (Thomas J., concurring in part and concurring in the judgment)).

explicitly and because a local preference is not understood in a racialized manner. On this understanding of what makes racial discrimination wrong, enhancing student performance is an adequate reason to prefer local applicants even when the policy works by lowering the number of minority applicants.

Perhaps a few more examples will make the approach I endorse clearer. Consider the practice of redlining, in which a mortgage lender refuses to lend money to home-buyers in specific neighborhoods. Suppose that this practice increases loan performance overall and does so because neighborhood is well-correlated with race which, in turn, is correlated with loan performance. I do not mean to assert that these correlations are accurate. Rather I use this example, and assume the truth of these suppositions, because redlining is a familiar example about which people's intuitions are fairly clear. In rejecting an approach that makes the actor's intentions relevant, I argue that we need not ask whether the lender, in adopting the neighborhood exclusions, intends to exclude buyers of color. Instead the relevant question is this: *Is enhancing loan performance an adequate reason to refuse loans to people buying houses in certain neighborhoods when this practice succeeds in improving loan performance by limiting the number loans offered to non-white home buyers?* Using this approach, the moral permissibility of redlining doesn't depend on the actor's intentions.

The objective approach I advocate here also does a better job of making sense of the sort of case in which it is unlikely that the actor specifically intends the proxy/target relationship that is supposed to be troubling. Consider exclusions of persons capable of becoming pregnant. Suppose an employer were to exclude persons capable of becoming pregnant from certain jobs because that trait (being capable of becoming pregnant) is well



correlated with susceptibility to injury in the work place. If persons who are capable of becoming pregnant are more likely than the average worker to be injured by the hazardous work environment, or if injured more likely to suffer a serious injury (miscarriage or impairment of the developing, fetus for example), then the proxy *person capable of becoming pregnant* may well be correlated with the target *likely to be injured in the workplace*. If we simply ask the question whether limiting the number and seriousness of workplace injuries is a sufficient justification for refusing to hire persons who are capable of becoming pregnant, we miss something important about this practice. In truth, it is hard to really miss it as we can't comprehend the classification *person's capable of becoming pregnant* without recognizing its overlap with the category *women*. But if we could, if we could somehow see it as just that – person's capable of becoming pregnant – we would miss something that is relevant to its permissibility. What we ought to ask, in considering its moral permissibility is this: *Is limiting the number and seriousness of workplace injuries a sufficient reason to refuse to hire persons capable of becoming pregnant when that proxy works by excluding women?* Whatever one's resolution of the question, the fact that it is women only who are capable of becoming pregnant is surely a central factor. And what's more: it is central whether or not the employer enacted the pregnancy exclusion in order to target women (which it most likely did not) or not.

To sum the conclusions of the argument of this section: the fact an actor intentionally uses a trait in a law or policy as a proxy for another more suspect trait – like race or sex – is not relevant to the moral permissibility of the law or policy. While it surely matters that the policy is successful or efficient because the proxy trait works via a

second proxy to reach its target, that fact is relevant whether it is intended or not. In reviewing a law, policy or decision, we need not ask whether the actor had the right intentions but instead whether the action is justified given the manner in which it operates.

### Part III.B The Purpose

The second sense of intention that might be relevant to the evaluation of whether a law, policy or decision constitutes wrongful discrimination is the *purpose for which* the law or policy is adopted. In this section, we will examine whether the fact that a law or policy is adopted *for a bad reason* is sufficient to render that law or policy impermissible. “Purpose” is often used interchangeably with “motive.” By “purpose,” I mean the end at which the actor aims. If intention is a function of the will – the actor’s purpose is the goal she deliberately seeks. Motive, by contrast, relates to the reason that actually moves the actor to act. The motive relates to what the actor desires rather than what she wills. But I don’t want to make too much of that distinction here. The view I endorse in this article is that the intentions of the actor – his consciously chosen ends – are not relevant to the moral permissibility of classificatory acts. While I also believe that his motives – the desire states that actually move him to action – are similarly irrelevant, I think that claim will follow from the argument presented here. In passages from cases cited below, these two terms – “purpose” and “motive” – are often used interchangeably and without careful attention to whether it is what the actor wills or what the actor desires that is thought to be central. As I want to argue that neither ought to be morally relevant to the

evaluation of whether laws or policies wrongfully discriminate, I ask the reader to treat them interchangeably in the passages below.

The view that a bad purpose is relevant to whether a law wrongfully discriminates seems to be the claim underlying Justice Brennan’s statement in *Moreno*, cited at the beginning of this article, that “if the constitutional conception of ‘equal protection of the laws’ means anything, it must at the very least mean that a bare congressional desire to harm a politically unpopular group cannot constitute a legitimate governmental interest.”<sup>53</sup> This Equal Protection minimum is reiterated in *Romer v. Evans*.<sup>54</sup> There the Court explains that one of the reasons in support of holding Colorado’s Amendment Two in violation of the Equal Protection Clause “is that laws of the kind now before us raise the inevitable inference that the disadvantage imposed is born of animosity toward the class of persons affected.”<sup>55</sup> The Court then goes on to cite the passage from *Moreno* above. But is bad purpose or motive – a bare desire to harm –relevant to whether a classification wrongfully discriminates or does it just seem so because the cases in which it is cited are over-determined? In other words, other wrong-making features are also present which confuse the issue.

In order to test the claim that invidious purpose or motive is relevant, we need to look at cases in which that bad motive is not accompanied by other wrong-making properties, else we won’t know if it is the motive that is really doing the work. Let’s begin with a silly example. Suppose a school in the process of hiring teachers posts the sign “No Martians need apply.”<sup>56</sup> Putting aside issues of standing – are there any

---

<sup>53</sup> 413 U.S. at 534.

<sup>54</sup> 517 U.S. 620 (1996).

<sup>55</sup> *Id.* at 634.

<sup>56</sup> This example was suggested to me by Julia Driver.

Martians to complain? – the School Board may intend to harm Martians. In that, they reveal themselves to be crazy, but do they also act wrongfully? It seems that at least something else is needed – either an effect or the expression of denigration.

More seriously, consider the following case. The orchestra of a large city is in the process of hiring new musicians. The musical director aims to keep as many Asian and Asian-Americans as possible out of the orchestra. Unfortunately for her, the orchestra auditions musicians behind a screen, to protect against this and other kinds of bias. In order to try to achieve her aim, the musical director picks musicians by counting passion more than technique, an approach she believes (mistakenly) will screen out as many Asian musicians as possible. As it turns out, the musical director bases her action on an inaccurate generalization. Asian musicians are not less passionate and more technically skilled than other musicians. As a result, the orchestra selected is one in which passionate musicians outnumber highly technical ones, but has the same racial composition (let us suppose) as would have been the case had the director adopted a different hiring policy. Ought her *bare* desire to harm to invalidate the hiring procedure?

My intuition, which I hope is shared, is that it ought not. If the desire or motive isn't actualized in any way, it is hard to see what makes the action wrong – though the desire surely reveals something unsavory about the musical director's character. But perhaps this is a bit too strong. There is something wrong about the actions of the orchestra director and the School Board, but what I want to argue here is that if there is something wrong about it, it is not the wrong of wrongful discrimination. Both the orchestra director and the School Board are no doubt subject to standards of practice that derive from their job descriptions – either explicitly or implicitly. The orchestra

director's job is to create the best orchestra, for example. If she also attempts to affect the racial composition of the orchestra, her actions are outside of her role, are an abuse of discretion, or something of that nature. This may well make her actions wrong, as violations of her role obligation. But this is not the wrong of wrongful discrimination. To see this point, we can imagine that the orchestra director was governed by no such role obligations. If so, the hiring procedure described above would not be wrong.

If bad purpose or motive isn't sufficient to render a law, policy or decision wrongful, perhaps it is *relevant* nonetheless. Doesn't a bad intention contribute to making a classification wrongful discrimination? The sort of case that tests this issue is one in which there is a bad intention and a bad effect but the effect, standing alone, would not render the classification wrongful. For example, suppose an employer adopts a policy of requiring employees to work 70-hours per week. Further suppose, not unreasonably, that this policy has the effect of rendering the workforce almost all male. Women, particularly women with children, are simply less willing to work the long hours that the employer demands. Suppose you believe, which is itself contestable (but let's leave this issue out for the moment) that this policy does not wrongfully discriminate against women or mothers. Does that judgment change if the employer adopts the policy deliberately in order to keep women out of his employ? At least at first blush, it does seem that this intention matters to our assessment of the wrongfulness of the policy. It seems to turn a permissible or ambiguous case into an instance of wrongful discrimination. Does this mean that the actor's intention really is relevant?

It does not – not when we examine the case more closely. If a long hours policy is permissible, despite its disproportionate negative effect on women, this must be

because it has some beneficial effects as well – it’s efficient. But there are other policies that the employer could adopt that would also produce efficiency, let’s say. This policy is chosen both because of its effect on productivity and its effect on the gender distribution of the work force. If there are several policies that would have increased productivity in more or less the same amounts but this one was adopted because of its gendered effects, then the employer really is using two criteria – sex and productivity – to choose which work-place policy to adopt. In that sense, it is not merely that he wills or wants fewer women, but that he chooses a policy of productivity/few women. It is not just his intention; it is what he *does*.

If the distinction I am drawing seems elusive, consider the following example. Suppose that an employer wishes to exclude Burmese employees. In order to do so, she adopts a policy of refusing to hire workers with the last name “Burman” or “Berman” due to the mistaken belief that all Burmese have last names of “Burman” or “Berman.”<sup>57</sup> Suppose that the only two Burmese applicants happened to be women who had taken their husband’s names and who were married to men with the last name Berman. In this case, we have an intent to exclude Burmese workers and the effect of excluding Burmese workers but we do not have a policy that discriminates on the basis of being Burmese.

In the long hours example, it made sense to say that sex was playing a role in the choice of policies by the employer. It wasn’t just that he intended or wished that sex be a factor, it was a factor in determining which productivity-enhancing policy he adopted. Here by contrast, the employer wills or desires that ethnicity play a role in her hiring decision, but it does not (except perhaps the excluding of Jewish employees as “Burman”

---

<sup>57</sup> This example was suggested to me by Mitchell Berman. While he used it to argue against my view, I have altered it somewhat – without his input – to make a slightly different point.

and “Berman” may be predominantly Jewish surnames – but that’s not the issue we’re considering). There is a coincidental effect of excluding Burmese job-seekers but we can’t really say that the policy wrongfully excludes Burmese because we can’t really say that the *policy* is one of excluding Burmese at all. The widely shared intuition that effect *alone* ought not to be enough to constitute wrongful discrimination grows out of this sort of fortuitous correlation between a policy and some consequence.

This argument harkens back to some of the discussion of Part II. Remember that there I argued that in order to know what the actual rule, policy or classification is, we may indeed have to go inside the head of the actor. So in cases of cognitive bias, race or sex or some other trait really is playing a role in the decision or policy at issue, notwithstanding the actor’s lack of intent to do so. Similarly, in the long hours example, sex really is one of the job criteria. We do need to go inside the actor’s head to determine this. But, and this is the key caveat, when we go inside, we do not look at what the actor intends but rather look at what she does. What she wills or desires are not relevant. We see that in the Burmese example. Even if the policy has the effect of excluding Burmese applicants, if the policy really doesn’t exclude people on the basis of Burmese ancestry, then this invidious intent and unfortunate effect do not render the policy wrongful.

## Conclusion

When we imagine posing the question *what makes classification wrongful* to the people affected by the laws, policies or decision at issue, we see that from their perspective it makes sense to be concerned with *how they are treated* rather than with passing judgment on state actors. To know how they are treated, we must focus on the

objective features of the laws and policies that affect them, rather than on the subjective intentions of the actors that adopt or enact these policies and laws. And Holmes would, most likely, agree. At the end of Lecture I of the Common Law, titled “Early Forms Liability,” in which the aphorism about dogs appears in the early part of his discussion, Holmes has the following to say:

It remains to be proved that, while the terminology of morals is still retained, and while the law does still and always, in a certain sense, measure legal liability by moral standards, it nevertheless, by the very necessity of its nature, is continually transmuting those moral standards into external or objective ones, from which the actual guilt of the party concerned is wholly eliminated.<sup>58</sup>

As the last part of his sentence makes clear, for Holmes, the “moral” refers to a judgment about the “actual guilt of the party.” In other words, Holmes too thinks that we often wrongfully confuse judgments about the moral character of the actor with judgments about the moral permissibility of the action. In assessing the wrongfulness of discrimination, it’s not the thought that counts.

---

<sup>58</sup> Holmes, *supra* note 1 at 33.