

**Intentional and Unintentional Discrimination:  
What Are They and What Makes Them Morally Different**

Forthcoming at *Journal of Moral Philosophy*; please cite the final version at:  
<https://doi.org/10.1163/17455243-20213430>

Rona Dinur  
Max Weber Programme, European University Institute, San Domenico di Fiesole, Italy  
[rdinur@gmail.com](mailto:rdinur@gmail.com)

**Abstract**

The distinction between intentional and unintentional discrimination is a prominent one in the literature and public discourse; intentional discriminatory actions are commonly considered particularly morally objectionable relative to unintentional discriminatory actions. Nevertheless, it remains unclear what the two types amount to, and what generates the moral difference between them. The paper develops philosophically-informed conceptualizations of the two types based on which the moral difference between them may be accounted for. On the suggested account, intentional discrimination is characterized by the agent viewing the content of an underlying discriminatory belief as a consideration that counts in favor of her action. This, it is argued, amounts to endorsing the discriminatory belief, which generates the particular moral severity of intentional discrimination.

**Keywords:** Discrimination; Equality; Implicit bias; Intentionality

1. Introduction

The distinction between intentional and unintentional discrimination is a prominent one in the discourse on discrimination and inequality. In American anti-discrimination jurisprudence, whether an alleged act of discrimination is intentional has long played a central role in determinations of its conformity with anti-discrimination norms.<sup>1</sup> The importance of the distinction, however, stretches far beyond the adjudication of claims of discrimination in legal venues; it is influential in political and everyday moral discourse, where allegations of intentionally discriminating against an individual or a group of people is

---

<sup>1</sup> A landmark case that introduced a rule along these lines is *Washington v. Davis*, 426 U.S. 229 (1976). While a legal determination that an action is intentionally discriminatory generally means that it violates anti-discrimination norms, the exact role that intentionality plays in the legal adjudication of claims of discrimination is complex, and several aspects of it remain obscure and controversial; and so are the relations between discriminatory actions being classified as intentional for legal purposes and them fitting within the legal categories of direct vs. indirect discrimination (or 'disparate treatment/impact discrimination'). For a recent overview see Aziz Z. Huq, 'What Is Discriminatory Intent?', *Cornell Law Review* 103 (2018), pp. 1211-1292. It seems evident that the prominence of the distinction in legal contexts draws, in part, on moral convictions about the particular wrongness of intentional discrimination, and thus the present discussion may clarify some aspects of anti-discrimination jurisprudence; however, due to the mentioned difficulties, drawing precise conclusions from the present discussion to legal contexts will require further work.

commonly taken to mark a particularly high degree of severity.<sup>2</sup> To give just a few recent examples, allegations of excessive use of force in policing have emphasized that such actions involve intentional discrimination against racial minorities; group-based travel and immigration bans have been accused of being intentionally discriminatory against the targeted groups; and opponents of diversity-based college admission policies have accused these institutions of intentionally discriminating against groups that are overrepresented in higher education. The notion of unintentional discrimination, on the other hand, has gained much attention in recent discussions across the social sciences and policy debates on the phenomenon of 'implicit bias': a variety of (possibly, unconscious) mental states that are thought to lead to unintentional discriminatory actions. It is commonly stated in these discussions that individual instances of unintentional discrimination are less morally abhorrent relative to more 'traditional,' intentional forms of discrimination (despite their significance in contributing to the persistence of inegalitarian societal patterns).<sup>3</sup>

The assumption commonly underlying this wide array of discussions is, then, that intentional discriminatory actions are particularly morally objectionable or severe, relative to unintentional ones. Despite the centrality of the distinction between the two and its significance in all of these debates, however, it remains unclear how the two types may be generally characterized, and how to demarcate the difference between them. In other words, the literature lacks precise conceptual analysis of the distinction. Since both intentional and unintentional discrimination are, presumably, thick moral concepts—both describing the nature of a particular phenomenon and morally condemning it in a way that is tied to its distinct characteristics—the absence of conceptual analysis leaves the reasons for the difference in moral severity between the two obscure as well. To my knowledge, these conceptual and ethical questions have not been examined in a way that fully integrates relevant theoretical knowledge developed in larger philosophical contexts.

The present paper thus seeks to develop conceptualizations of intentional and unintentional discrimination that clearly demarcate the conceptual difference between them, based on which the putative moral difference between the two types may be plausibly accounted for. Considering that not much attention has been devoted to this endeavor so far, the discussion here pursues this task in an exploratory manner. Specifically, it aims at developing conceptualizations that significantly contribute to our understanding of these phenomena and adequately accommodate central, clear-cut examples, but not necessarily at accommodating all possible cases or presenting necessary and sufficient

---

<sup>2</sup> For philosophical overviews see Andrew Altman, 'Discrimination,' in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (2016), sec. 2.1. <<http://plato.stanford.edu/archives/win2016/entries/discrimination>>; Kasper Lippert-Rasmussen, *Born Free and Equal? A Philosophical Inquiry into the Nature of Discrimination* (Oxford: Oxford University Press, 2013), pp. 55-61, 111-12.

<sup>3</sup> See, e.g., B. Keith Payne and Heidi A. Vuletic, 'Policy Insights from Advances in Implicit Bias Research,' *Policy Insights from the Behavioral and Brain Sciences* 5 (2018), p. 50. Similar thoughts are present in recent discussions on 'systemic' or 'structural' discrimination. See, e.g., Roland L. Craig, *Systemic Discrimination in Employment and the Promotion of Ethnic Equality* (Leiden; Boston: M. Nijhoff, 2006), pp. 93-94.

conditions for intentional and unintentional discrimination.<sup>4</sup> Further, in order to keep the discussion clear and manageable—among other things, by avoiding complications accompanying the appraisal of discriminatory actions performed by collective agents, and confabulating factors that are often present where they are performed by public institutions, state officials and politicians—the discussion focuses only on individual, identifiable discriminatory actions performed by individuals in their private capacity, in the course of their interpersonal interactions. As will become clear in Section 5, the discussion is friendly to the view that these actions exhibit the normative phenomenon at issue in its most vivid form; however, its conclusions may be extended to discriminatory actions performed in other contexts and by other types of agents, depending on one's position on the theoretical issues which these may raise.

On the suggested account, intentional discrimination consists of a discriminatory behavior that satisfies the following conditions: it is intentional under a description with a discriminatory structure, along the lines of 'differentiating between socially salient groups X and Y'; and involves the agent acting based on a motivating reason making reference to the content of a discriminatory belief, e.g., 'members of group X are inferior (evil, criminal, etc.).' By contrast, in unintentional discrimination, an underlying discriminatory mental state causally influences the agent's behavior,<sup>5</sup> making her act in a discriminatory manner consistent with a discriminatory proposition; but she does not incorporate this discriminatory content into her motivating reason. This, I argue, generates the moral difference between the two types: in acting with the motivational structure characteristic of intentional discrimination the agent endorses her underlying discriminatory belief, thus acting with a particularly objectionable attitude. Considering that discriminatory attitudes (partly) generate the moral objection to discriminatory actions, this heightened severity is what makes intentional discrimination particularly objectionable.

Before proceeding, it is important to clarify that the paper is concerned with the moral status of discriminatory actions, and not with the agent's blameworthiness or moral responsibility for performing them or holding discriminatory mental states.<sup>6</sup> I also separate the questions discussed here from evaluations of the agent's character (as a racist, sexist, etc.). While these three distinct moral evaluations seem interrelated in this particular context, considering the discussions' exploratory character (and the nascent nature of discussions on these related topics), it seems premature to attempt to characterize such possible relationship in detail. I therefore take the methodological step of assuming that the ethical questions addressed here can be usefully analyzed in isolation from these other questions.

---

<sup>4</sup> Similarly, I cannot defend the suggested account against possible alternative conceptualizations, which may be developed based on some common usages of the term 'intentional discrimination.' As noted in Section 2 (fn 16), however, this account possibly subsumes these conceptualizations.

<sup>5</sup> Throughout, when I say that mental states causally influence the agent's action or behavior, I mean that they so in the particular manner in which mental states causally influence actions.

<sup>6</sup> I do assume, however, that these actions and mental states are attributable to the agent—making the agent responsible for them in this minimal sense—despite possibly being (in some sense) unconscious, and out of the agent's control. This approach roughly aligns with Michael S. Brownstein's in *The Implicit Mind: Cognitive Architecture, the Self, and Ethics* (New York: Oxford University Press, 2018), ch. 4-5.

The discussion proceeds as follows. Section 2 presents some background about the nature of discriminatory actions. Conceptualizations of intentional and unintentional discrimination are developed in Sections 3 and 4, respectively; the conceptual difference between the two is closely examined in these two sections. Section 5 discusses the moral difference between the two types. It first discusses the reasons why mental states or attitudes are important in generating the moral status of actions that are similar to discriminatory actions in important respects; and then shows that there are particular reasons to think that whether discriminatory beliefs are intentionally acted on—that is, by incorporating their content into one's motivating reasons—matters in evaluating the severity of discriminatory actions in particular, as this generates the mentioned endorsement. Section 6 concludes.

## 2. Discriminatory Actions: Background

As mentioned, the discussion focuses on individual, identifiable discriminatory actions performed by individuals in their private capacity, in the course of their interpersonal interactions. Here are some central, clear-cut examples of intentional discriminatory actions of this type:

*Small Shop.*<sup>7</sup> Upon spotting a person whom she believes to be of Roma descent, a business owner prevents her from entering her small shop, shouting 'Roma out!'

*Prison.*<sup>8</sup> A warden in apartheid South Africa gives shorts to Black inmates, and long pants to White inmates. She believes that wearing shorts is humiliating; regardless, allowing inmates to wear shorts provides them with a material advantage.

*Job Interview.* When conducting job interviews, a small business owner is deliberately less friendly towards female than male candidates. In making hiring decisions, she sometimes rejects female candidates on the spot, without considering their professional credentials.

Unintentional discriminatory actions, by contrast, are usually considered to stem from the psychological phenomenon of implicit bias. Here are central examples:

*Unfriendly Behavior.*<sup>9</sup> Due to her (possibly, unconscious) association of the word 'black' with concepts with negative valence (e.g., 'bad'), an agent displays less friendly behaviors towards African-American interlocutors, relative to White-American ones.

---

<sup>7</sup> Example from Altman, 'Discrimination.'

<sup>8</sup> This example is revised from Deborah Hellman, *When Is Discrimination Wrong?* (Cambridge: Harvard University Press, 2008), p. 5.

<sup>9</sup> All examples of unintentional discrimination discussed herein are based on psychological research. See generally, Anthony G. Greenwald and Mahzarin R. Banaji, 'Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes,' *Psychological Review* 102 (1995), pp. 4-27; Russell H. Fazio, Joni R. Jackson, Bridget C. Dunton, and Carol J. Williams, 'Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: a Bona Fide Pipeline?,' *Journal of Personality and Social Psychology* 69 (1995), pp. 1013-27. For this particular behavior, see John F. Dovidio Kerry Kawakami, and Samuel L. Gaertner, 'Implicit and Explicit Prejudice and Interracial Interaction,' *Journal of Personality and Social Psychology* 82 (2002), pp. 62-68.

*Shooting Bias*.<sup>10</sup> Due to her (possibly, unconscious) association of African-Americans and concepts related to crime and aggression (or beliefs with corresponding discriminatory propositional content), a police officer perceives an ambiguous object as a gun when held by a Black suspect, while perceiving the same item as a neutral object when held by a White suspect. Upon spotting a black-skinned suspect on a dimly lit street, she makes a split-second decision to pull out her gun and shoot the suspect (while she would not do the same if the suspect were white-skinned).

To characterize the basic nature of both intentional and unintentional discriminatory actions, I utilize (a slightly revised version of) Andrew Altman's and Kasper Lippert-Rasmussen's conceptual analysis,<sup>11</sup> according to which they consist of *disadvantageous differential treatment of an individual (or a group of people), in a way that corresponds to their membership in a socially salient group*. 'Socially salient group' refers here to a group that is important in regard to a wide range of social interactions in the relevant society. For simplicity, and considering that the particular nature of the group involved is immaterial for the paper's main claims, I focus on cases involving race, ethnicity, or gender.

In line with common and robust intuitions—prevalent in both the philosophical and legal literature, as well as public discourse—I assume that such actions intrinsically violate equality, that is, in a way tied to their distinct discriminatory character, and regardless of their contingent consequences (including consequences such as exacerbating inequalitarian societal patterns).<sup>12</sup> This can be straightforwardly supported by noting that cases such as *Prison* and *Unfriendly Behavior* do not involve any objectionable consequences (such as harming the discriminatee, materially disadvantaging them in any way, or subjecting them to an unjust distribution of resources), while undoubtedly exhibiting the core wrong associated with discrimination. Similarly, where the objectionable consequences that are presumably present in cases such as *Job Interview* and *Shooting Bias* are assumed away, this does not seem to affect the objection associated with the action's discriminatory features. Note that this entails that the moral difference which putatively stems from whether or not such actions are intentional under a description making reference to their discriminatory features—for which the present discussion aims to account—is distinct and independent from a difference in their moral status that might be generated by whether or not they are intentional under a description making reference to their objectionable consequences (in cases involving such consequences).<sup>13</sup> For simplicity, then, I assume away any such objectionable consequences throughout the discussion.

More particularly, what seems to be the wrong-making features of these actions are the agent's discriminatory behavior—the differential treatment which corresponds to the victim's group identity—combined with her underlying discriminatory mental state(s). This thought too is supported by widespread and persistent intuitions associated with the key

---

<sup>10</sup> B. Keith Payne, 'Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon,' *Journal of Personality and Social Psychology* 81 (2001), pp. 181-92.

<sup>11</sup> Altman, 'Discrimination,' sec. 1; Kasper Lippert-Rasmussen, *Born Free and Equal?*, pp. 13-46.

<sup>12</sup> Kasper Lippert-Rasmussen, 'Discrimination and Equality,' in Andrei Marmor (ed.) *The Routledge Companion to Philosophy of Law* (New York and London: Routledge, 2012), pp. 569-83.

<sup>13</sup> In that sense, the ethical question at issue here is importantly different from the one addressed in debates over the Doctrine of Double Effect. See the discussion in Section 5A (and fn 39).

examples presented above, along with some central accounts of discrimination offered in the literature.<sup>14</sup> Thus, for example, while it seems unobjectionable for a warden to supply different types of clothing to inmates belonging to different socially salient groups based on her beliefs about their different physical needs, it would be objectionable to do so based on a belief that some prisoners belong to a group deserving to be marked as inferior. Similarly, while it would be permissible for a homeowner not to sell her house based on some arbitrary or whimsical reason, it would be objectionable to do so in a way influenced by group-based hatred.

I discuss the justification for this view in more detail in Section 5; at this point, note that discriminatory mental states may include beliefs with discriminatory or inegalitarian content (e.g., that 'All members of group X are inferior, bad, criminals'), negative emotions such as hate or contempt directed at the discriminated group, and discriminatory conceptual associations (e.g., associating the word 'Black' with the word 'bad' or 'angry', while 'White' is associated with 'good' or 'calm'). Since it is not my goal here to settle questions about which contents (of mental states) are adequately characterized as objectionably discriminatory (for instance, whether all group-based generalizations are objectionable, or there are 'benign' generalizations<sup>15</sup>), I use the term 'discriminatory/inegalitarian mental state' as a placeholder to denote any mental state whose content is objectionably discriminatory or inegalitarian. If one disputes a particular example, she may replace it with a more clearly objectionable one, e.g., mental states attributing degrading or dehumanizing traits to the discriminated group, or a belief that the group is inferior in some important respect.

Reflecting the nature of the clear-cut examples of intentional discrimination in particular, I take such actions to be underlain by a (conscious or unconscious) discriminatory *belief* (while they may or may not be underlain by other discriminatory mental states as well). It is plausible that in *Prison*, for instance, the discriminator believes that Black people are inferior to Whites in some important respect (and therefore deserving of humiliation); in *Small Shop*, the discriminator seems to believe something along the lines of 'People of Roma descent are [negative trait] and should not be associated with'; and in *Job Interview*, that 'Women are unsuitable for the workplace.'<sup>16</sup>As will become clear, the nature of the mental state appropriately attributed to agents performing unintentional discriminatory actions is a matter of debate; I discuss this point in Section 4.

---

<sup>14</sup> This view aligns most naturally with mental-state accounts of discrimination. See, e.g., Larry Alexander's (now retracted) account in 'What Makes Wrongful Discrimination Wrong? Biases, Preferences, Stereotypes, and Proxies,' *University of Pennsylvania Law Review* 141 (1992), pp. 149-219; and overview at Lippert-Rasmussen, *Born Free and Equal?*, pp. 103-128. However, it seems compatible with any account that lends *some* role to mental states in generating the wrongness of discriminatory actions, including, e.g., Benjamin Eidelson's in *Discrimination and Disrespect* (Oxford: Oxford University Press, 2015).

<sup>15</sup> See Erin Beeghly, 'What Is a Stereotype? What Is Stereotyping?,' *Hypatia* 30 (2015), pp. 675-691; Lawrence Blum, 'Stereotypes and Stereotyping: A Moral Analysis,' *Philosophical Papers* 33 (2004), pp. 251-289.

<sup>16</sup> Some notions of intentional discrimination associate it with hate, animosity, or a desire to exclude the group. It is plausible that in such cases the action is *also* underlain by a discriminatory belief, and thus the account developed here might subsume these general notions.

### 3. Intentional Discrimination

The discussion in this section oscillates between close examination of the cases and laying out the relevant theoretical background about the motivational phenomena that they involve—including intentional action, intentions, and motivating reasons. From the vast literature on these topics, I adopt a set of theoretical positions that, while not universally accepted, are prominent in the relevant debates and seem particularly useful in developing the requisite conceptualizations. This, of course, requires simplification and ignoring difficulties with these positions that have been discussed in the literature. Note, however, that the cases examined here often do not exhibit the features based on which challenges to these positions have been mounted—and so it is possible that even if one holds theoretical positions that diverge from the ones adopted here, this would not have significant implications for the discussion.<sup>17</sup>

#### A. Intentional Action, Intentions, and Discriminatory Actions

I employ certain theoretical assumptions on the nature of intentional action and intentions. First, I use a coarse-grained individuation of action. To use Davidson's famous example, if by flipping a switch I turn on the light and also alert a burglar—all of these constitute a single action; the action would be intentional under a certain description, for instance, 'turning on the light,' but not 'alerting the burglar.'<sup>18</sup> I further assume that there are behaviors or bodily movements which constitute genuine actions—that is, that are not *mere* bodily movements or reflexes—which are not, on the other hand, cases of fully intentional action. That is, I allow for an in-between category of *unintentional action*;<sup>19</sup> as this might be controversial, Section 4 briefly discusses the justification for this view, and illustrates that there are good reasons to suppose that central cases of unintentional discrimination—such as *Shooting Bias*—fall within this category. With respect to actions that *are* intentional, I assume that they may reflect varying levels of agency or autonomy. For instance, where an agent consciously acts on a motivating reason (as will become clear, by 'motivating reasons' I mean considerations which the agent views as counting in favor of her action), her action may reflect a higher level of agency or autonomy, relative to an action where she acts (intentionally) on unconscious motivating reasons or for no reason at all. Such an elevated

---

<sup>17</sup> For instance, Bratman's famous objection to the 'Simple View' (the view employed here) about the relation between intentional action and intentions has been developed based on problems posed for it by actions involving complex sets of intentions, which are not typically present in cases discussed here. See Michael Bratman, *Intention, Plans, and Practical Reason* (Cambridge: Harvard University Press, 1987), ch. 8-9.

<sup>18</sup> Donald Davison, 'Actions, Reasons, and Causes,' *The Journal of Philosophy* 60 (1963), pp. 685-700; G. E. M. Anscombe, *Intention*. 2nd edn (Ithaca: Cornell University Press, 1957).

<sup>19</sup> Tamar Szabó Gendler, 'Between Reason and Reflex: Response to Commentators,' *Analysis* 72 (2012), pp. 799–811; Bence Nanay, *Between Perception and Action* (Oxford: Oxford University Press, 2013), pp. 81-86.

level of agency or autonomy is not required, however, for intentional action; some actions are adequately characterized as intentional despite lacking these features.<sup>20</sup>

I also assume certain things about the relation between intentional action and the agent's present-directed intentions, or the intentions she holds with respect to her present action<sup>21</sup> (as distinct from intentions for the future; the latter are irrelevant for the discussion here). First, I take present-directed intentions to be a distinct type of mental state, involving an attitude of resoluteness or settledness towards a certain course of action;<sup>22</sup> second, I assume that intentionally A-ing entails intending to A, or holding a present-directed intention to A; and third, that the description under which a given action is intentional (roughly) corresponds to the content of the agent's present-directed intention.<sup>23</sup> Thus, where an action is intentional under the description 'turning on the light,' this means that the agent acts with a present-directed intention to 'turn on the light.'

Some additional clarifications are needed considering some (implicit or explicit) assumptions common in the discourse on discrimination in particular. First, present-directed intentions do not have to be formed prior to a given intentional action; they may accompany an intentional action which the agent spontaneously embarks on. Consider the business owner in *Small Shop*: it is plausible she had not formed an intention prior to performing her action; rather, she may have spontaneously responded to seeing the victim by acting in a discriminatory manner. Her action seems intentional nevertheless.<sup>24</sup> Relatedly, it is important to highlight that performing an intentional action does not require that the agent be conscious, or consciously aware, of performing it (at least on some ways of understanding the nature of such consciousness). Similarly, acting with a certain intention does not necessitate consciously entertaining it, either prior to or at the point of acting (and similar things can be said about the agent's motivating reasons, discussed below). Consider again the discriminator in *Small Shop*: she may have not consciously entertained her intention prior to or at the point of acting, nor may she have been consciously aware of performing an intentional action—perhaps because she acts out of habit, or without attending to her action when performing it; her action seems intentional nevertheless.<sup>25</sup>

---

<sup>20</sup> James David Velleman, *The Possibility of Practical Reason* (Oxford and New York: Clarendon, 2000), pp. 1–31; Alfred R. Mele, *Motivation and Agency* (New York: Oxford University Press, 2003), pp. 215–32.

<sup>21</sup> Or the agent's 'intention-in-action'; John R. Searle, *Intentionality: an Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983).

<sup>22</sup> Hugh J. McCann, 'Settled Objectives and Rational Constraints,' *American Philosophical Quarterly* 28 (1991), p. 26.

<sup>23</sup> Frederick Adams, 'Intention and Intentional Action: The Simple View,' *Mind & Language* 1 (1986), pp. 281–301.

<sup>24</sup> Searle, *Intentionality*, pp. 84–85.

<sup>25</sup> Anscombe, *Intention*, and Kieran Setiya, 'Knowledge of Intention,' in Anton Ford, Jennifer Hornsby, and Frederick Stoutland (eds.), *Essays on Anscombe's Intention* (Cambridge: Harvard University Press, 2011), pp. 170–197; Bill Pollard, 'Explaining Actions with Habits,' *American Philosophical Quarterly* 43 (2006), pp. 62–63. A common thought in both legal and psychological literatures associates intentionality with consciousness in the context of discrimination, but often it is not clear what is meant by 'conscious discrimination' in these discussions. As intentionality is not commonly considered to necessitate consciousness in philosophical discussions, we should remain open to the possibility that there are intentional discriminatory actions which are unconscious in some sense.



Let us examine intentional discriminatory actions in light of these theoretical premises. Discriminatory actions (both intentional and unintentional) are characterized by a distinct structure or pattern: a single action consists of several simpler or more basic actions, which diverge in a way corresponding to the group identity of the people towards which they are directed (e.g., behaving in a friendly manner towards male job candidates and in an unfriendly manner towards female candidates). Presumably, then, for a given discriminatory behavior to constitute intentional discrimination, it has to be intentional under a description incorporating this particular feature. To wit, it is not enough that each of the actions constituting the discriminatory pattern is intentional on its own; rather, the differentiation between the simpler actions and its correspondence to the group identity of the people towards which these simpler actions are directed has to be intentional.

Hence, for discriminatory behavior to constitute intentional discrimination, it has to be intentional under a description such as 'differentiating between inmates in the distribution of pants, in a way corresponding to whether they are Black or White,' or 'denying entry to people of Roma descent, while allowing entry to people of non-Roma descent.' Since the nature of the actions of which the discriminatory pattern is composed is not itself important, the description under which the action is intentional must have a general discriminatory structure—such as 'treating socially salient groups X and Y differently,' or 'treating people differently in a way corresponding to their group identity,' or any essentially similar description. In line with the mentioned theoretical premises, the content of the agent's intention in such a scenario would correspond to this description, and have a generally discriminatory structure.

Understanding the functional role of intentions (along with that of motivating reasons, discussed below) will be important in accounting for the putative moral difference between intentional and unintentional discrimination. As mentioned, I take intentions to consist of a distinct attitude of resoluteness towards a certain course of action; in this sense, they involve a practical commitment to this course of action. As for their functional role, it is often thought to involve providing a settled objective for an (intentional) action, and controlling or sustaining its execution in light of that objective.<sup>26</sup> It can be said, then, that in acting with a certain intention, the agent is practically committed to controlling her course of action in light of the settled objective provided by her intention. In the case just mentioned, for instance, the warden is practically committed to 'differentiating between inmates in the distribution of pants, in a way that corresponds to whether they are Black or White.' I return to this point in Section 5.

I have argued that intentional discrimination requires that the agent's discriminatory behavior be intentional under a description with a discriminatory structure. Clearly, however, this is not enough; such an action may be entirely unobjectionable. This possibility is illustrated by a case where an agent intentionally provides different types of clothing to members of different social groups, where her underlying reasoning makes reference to the different needs of these groups; such an action would be 'intentionally discriminatory' only in a non-moralized sense. This is not a surprising conclusion: I started the discussion by

---

<sup>26</sup> McCann, 'Settled Objectives and Rational Constraints,' p. 26; Robert Audi, 'Intention, Cognitive Commitment, and Planning,' *Synthese* 86 (1991), pp. 363-64.

highlighting that the wrong-making features of discriminatory actions are the discriminatory behavior *in conjunction with* the underlying inegalitarian mental state—an inegalitarian belief in the case of intentional discrimination. Presumably, then, wrongful intentional discrimination involves a motivational structure that incorporates this belief in some appropriate way. In the next subsection, I argue that this is best explicated as a requirement that the agent act on a motivating reason incorporating the content of her underlying discriminatory belief.

## B. Intentional Discrimination and Motivating Reasons

Extensive discussions are devoted in the literature to motivating reasons, and there is much divergence in the terminology which authors employ, and the substantive positions defended about the nature of motivating reasons and the role they play in action. Some clarifications of terminology, and about larger theoretical issues raised in these debates, are needed to effectively present the theoretical assumptions employed here.<sup>27</sup>

First, I use the term 'motivating reasons' to refer to considerations which the agent views as counting in favor of her action, and in light of which she acts.<sup>28</sup> It is easy to identify the content of a motivating reason by examining what the agent would say if asked why she acted the way she did; presumably, even if the agent is not conscious of her motivating reasons at the time of acting, they are in principle available to her consciousness, and their content would be conveyed by the answer to that question. I assume that motivating reasons are conceptually distinct, and may diverge in their content in a given action from 'normative reasons'—the latter are considerations that *actually* count in favor of the action (if there are such). Presumably, in many cases of wrongful intentional discrimination the consideration the agent views as counting in favor of her action—whose content consists, as discussed below, of the content of her discriminatory belief—does not, of course, constitute a normative reason at the same time.

I assume that motivating reasons, or the content of an agent's motivating reasons, consist of the content of her (possibly false) beliefs, i.e., putative or believed facts (or, alternatively, propositions or the truth-makers of propositions); this is contrary to some views on which motivating reasons consist of beliefs. These believed facts may not obtain in reality, as is the case with the content of many discriminatory beliefs.<sup>29</sup> Relatedly, the literature sometimes talks of 'explanatory reasons' (or 'reason explanations'); I use these

---

<sup>27</sup> This overview, along with the particular terminology and substantive positions employed here, is based on María Alvarez's discussion in *Kinds of Reasons: An Essay in The Philosophy of Action* (Oxford: Oxford University Press, 2010) and 'Reasons for Action: Justification, Motivation, Explanation,' in Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (2017), <https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/>. I also assume that acting based on motivating reasons entails that the action is intentional (but not the other way around). See María Alvarez, 'Acting Intentionally and Acting for a Reason,' *Inquiry* 52 (2009), pp. 293–305.

<sup>28</sup> Other terms used to denote roughly the same phenomenon include 'the agent's reason' and 'agential reasons.'

<sup>29</sup> Alvarez, 'Reasons for Action,' sec. 3.1; *Kinds of Reasons*, pp. 124–64. As will become clear, I do not use 'motivating reasons' as a factive term in a sense implying that these putative facts play a causal role in the agent's action.

terms to refer to mental states that are causally responsible for the action and may provide a causal explanation of it, and assume that there is a conceptual distinction between these and motivating reasons (and between them and normative reasons). At a minimum, where an agent acts based on a motivating reason making reference to a (putative) fact that does not actually obtain, her motivating reason diverges from the explanatory reason of the action: the former consists of this putative fact, while the latter consists of her false belief representing the relevant fact as obtaining. I assume, then, that when an agent acts on a motivating reason, the belief that represents the relevant fact as obtaining must be part of what explains her action in causal terms (while there may also be other mental states that are 'reason explanations' of her action).

With these clarifications in mind, we can turn to understanding the role of motivating reasons in the conceptualization of wrongful intentional discrimination. As already noted (and further discussed in Section 5), intentional discriminatory actions are underlain by a discriminatory belief, which is jointly constitutive—with the group-based differential treatment—of their wrongness. It therefore seems that an adequate conceptualization of intentional discrimination has to incorporate the discriminatory belief in the agent's motivational structure in some appropriate way. We have also seen that the content of one's motivating reasons makes reference to the content of one's beliefs. Hence, a natural way in which a certain belief may be incorporated into one's motivational structure is by incorporating its content into one's motivating reason. It can be said, then, that wrongful intentional discrimination requires both that (1) the discriminatory behavior be intentional under a description with a discriminatory structure, and (2) the agent's motivating reason makes reference to the content of her underlying discriminatory belief.

This suggestion makes theoretical sense, and seems to adequately accommodate the central, clear-cut examples of intentional discrimination. For instance, in *Small Shop*, it is plausible that the agent incorporates the content of her discriminatory belief—presumably, 'People of Roma descent are [negative trait] and should not be associated with'—in her motivating reason. If asked why she acted the way she did, she would presumably say, 'I did not let that person into my shop [while letting in people of non-Roma descent] because people of Roma descent are [negative trait].' Similar things may be said about the agents in *Prison* and *Job interview*: they would be disposed to present the content of their underlying discriminatory belief if asked why they acted the way they did, that is, if asked to point to the content of their motivating reason.

As further discussed in Section 5, the functional role of motivating reasons is significant in generating the moral difference between intentional and unintentional discrimination. However, beyond the general thought that motivating reasons *guide* the agent in performing her action or executing her intention, not much attention has been devoted to this functional role in the literature.<sup>30</sup> Here, I tentatively suggest—without purporting to fully explicate or defend this claim—that this common thought is plausibly

---

<sup>30</sup> See brief discussions in Dustin Locke, 'Knowledge, Explanation, and Motivating Reasons,' *American Philosophical Quarterly* 52 (2015), pp. 215–232; Raimo Tuomela, 'Motivating Reasons for Action,' in Mark Timmons, John Greco and Alfred R. Mele (eds.), *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi* (New York: Oxford University Press, 2007), pp. 178–84.

explicated along the following lines. While the agent's intention controls her overarching course of action in light of its objective, her motivating reason(s) determine the nature of the subsidiary (or fine-grained) actions of which this overarching course of action is composed.<sup>31</sup> Thus, in executing her intentional action, an agent is disposed to adjust these subsidiary actions so that they would match the content of her motivating reason(s) in some appropriate way. This means that while an agent's overarching course of action would be identical in two actions where she acts with an identical intention, her subsidiary actions would diverge if she acts with diverging motivating reasons.

This divergence seems particularly important where discriminatory actions are concerned. Consider two agents acting with an identical intention, involving, for instance, the distribution of different types of clothing to individuals belonging to different socially salient groups. The motivating reason of Agent A appeals to the proposition 'Group X is inferior to Group Y, and should be humiliated,' while the motivating reason of Agent B appeals to the putative fact that 'Group X and Group Y have different physical needs when it comes to clothing.' The overarching course of action in these two scenarios would be identical, and involve distributing different types of clothing in a way that corresponds to the group identity of the recipients. However, in executing her intention, Agent A would pick items of clothing considered humiliating (and give them to Group X), while Agent B would choose items of clothing that are accommodating of each group's needs. Similarly, the two agents would change these subsidiary actions in different ways upon finding out that certain pieces of clothing are suitable for Group X's physical needs—this would influence the subsidiary actions of Agent B, but not of Agent A.

Following the comments about the practical commitment embedded in acting with a certain intention, it may accordingly be said that the content of one's motivating reasons determines the content of one's practical commitment, when the latter is characterized in a fine-grained manner, that is, one that considers not only the agent's overall course of action, but also her fine-grained, subsidiary actions. Thus, where Agent A's fine-grained practical commitment is along the lines of 'Treating inferior Group X and superior Group Y differently in the distribution of clothing,' Agent B is practically committed to 'differentiating between Group X, which has certain physical needs, and Group Y, which has different physical needs.' More generally, in acting intentionally with a motivating reason, an agent is practically committed to executing her intention in a way which appropriately matches the content of her motivating reason(s). I return to this point in Section 5.

#### 4. Unintentional Discrimination

Unintentional discriminatory actions—as in *Unfriendly Behavior* and *Shooting Bias*—are commonly considered to stem from the psychological phenomenon of implicit bias. A

---

<sup>31</sup> See discussion of subsidiary actions in Markos Valaris, 'The Instrumental Structure of Actions,' *The Philosophical Quarterly* 65 (2015), pp. 64-70, 76-77.

preliminary challenge in conceptualizing unintentional discrimination is tied to the fact that it is still not clear what these implicit biases consist of. In line with central positions in recent discussions, I assume that implicit biases are mental states, and remain neutral as to whether they are conceptual associations, beliefs, or another type of mental state (such as 'aliefs').<sup>32</sup>

This assumption does not seem problematic, as it is inconsequential for our purposes. Since the present discussion is concerned with the moral status of discriminatory actions—and not, as in recent literature on implicit bias, with the nature and moral evaluation of the implicit biases themselves, or with the agent's moral responsibility for harboring them—what seems important is the role which implicit biases play in generating unintentional discriminatory actions. Here there is not much controversy: regardless of the type of mental state they are, implicit biases are widely thought to *causally influence* the agent to act in a discriminatory manner—particularly one consistent with believing a discriminatory proposition corresponding to their content. Thus, whether or not the agent in *Unfriendly Behavior* believes that 'Black people are bad' or alternatively only associates the concept 'Black' with 'Bad,' she is causally influenced by her implicit bias to act in a discriminatory manner consistent with believing that 'Black people are bad.'<sup>33</sup> It seems consensual, on the other hand, that in performing unintentional discriminatory actions the agent does not view this discriminatory propositional content as a consideration counting in favor of her action, and does not guide her action so it would match this content in the manner characteristic of acting based on motivating reasons. Here too, a close examination of the cases is illustrative.

A plausible way of understanding *Unfriendly Behavior* is the following. The agent in this case may intentionally treat her conversational counterparts differently—for instance, she may smile more or less at or stand closer or farther away from an interlocutor—based on how pleasant she feels around them, or on whether or not they seem to display positive emotions. Alternatively, she may not intend to differentiate between interlocutors at all. Either way, due to her implicit bias she consistently feels less positive emotion around Black people, or tends to perceive them as exhibiting more negative emotions—which leads her to display fewer friendly behaviors around members of the group.<sup>34</sup> Thus, she is causally influenced by her implicit bias to act in a discriminatory manner consistent with believing a

---

<sup>32</sup> See discussion in Michael Brownstein and Jennifer Saul, 'Introduction,' in *Implicit Bias and Philosophy*, pp. 1–9 (implicit biases are conceptual associations); Tamar Szabó Gendler, 'On the Epistemic Costs of Implicit Bias,' *Philosophical Studies* 156 (2011), pp. 33–63 (associative 'aliefs'); Eric Mandelbaum, 'Attitude, Inference, Association: On the Propositional Structure of Implicit Bias,' *Noûs* 50 (2016), pp. 629–58 (beliefs or belief-like mental states); Neil Levy, 'Neither Fish nor Fowl: Implicit Attitudes as Patchy Endorsements,' *Noûs* 49 (2015), pp. 800–23 (a *sui generis* mental state). I also remain neutral on whether implicit biases are unconscious (and in what sense).

<sup>33</sup> This assumption seems to motivate debates on whether agents are adequately attributed discriminatory beliefs based on such unintentional actions. See, e.g., Eric Schwitzgebel, 'Acting Contrary to Our Professed Beliefs or the Gulf Between Occurrent Judgment and Dispositional Belief,' *Pacific Philosophical Quarterly* 91 (2010), pp. 531–553.

<sup>34</sup> See Arianna Bagnis, Alessia Celeghin, Cristina Onesta Mosso, and Marco Tamietto, 'Toward an Integrative Science of Social Vision in Intergroup Bias,' *Neuroscience and Biobehavioral Reviews* 102 (2019), pp. 318–326; and discussion in Alex Madva and Michael Brownstein, 'Stereotypes, Prejudice, and the Taxonomy of the Implicit Social Mind,' *Noûs* 52 (September 2018), pp. 611–644.

discriminatory proposition such as 'Black people are bad,' but does not see this content as a consideration counting in favor of her action.

*Shooting Bias* seems like a different type of case. As mentioned above, some behaviors seem to constitute genuine actions, but at the same time not to constitute a fully intentional action (some call these semi-, or quasi-intentional actions). Gendler claims that they occupy a middle ground 'between reason and reflex.' The category is unique because unlike reflexes or mere behaviors, the representational content of the underlying mental state plays a significant role in initiating the action and determining its nature; however, unlike fully intentional actions, it is processed only in a shallow manner: one that does not fully involve the agent's reasoning capacities.<sup>35</sup> *Shooting Bias* seems to fall into this category, and it too can therefore be described as a case where the agent is causally influenced by her underlying mental state to act in a discriminatory manner consistent with believing, e.g., that 'Black people are criminals'—but where she does not incorporate this content into her motivating reason.

Both of these central examples of unintentional discrimination, then, involve a discriminatory action—one that is either intentional under a non-discriminatory description or quasi-intentional—that does not satisfy the *second* condition specified above: in both, the agent does not incorporate the propositional content of her underlying discriminatory mental state in her motivating reason. Note that while these cases also do not satisfy the *first* condition—both are not intentional (or fully intentional) under a description with a discriminatory structure—it is possible in principle to think of an objectionable discriminatory action, i.e., one where the differential treatment is underlain by a discriminatory mental state, which satisfies the first but not the second condition. Suppose, for instance, that an agent intentionally differentiates between men and women in job interviews, where her motivating reason refers to the (putative) fact that her business needs more male employees (for some legitimate reason). In performing her action, however, she is partly causally influenced by her unconscious association of 'women' with 'home life,' making her rate women lower on measures of work-related skills. Such a case seems intuitively objectionable only as an instance of unintentional (but not intentional) discrimination.

The opposite scenario seems much less likely. This would involve a discriminatory behavior underlain by a discriminatory mental state, where the agent incorporates the discriminatory content of the mental state in her motivating reason (which necessarily, on the theoretical premises adopted here, makes the action fully intentional)—but where the action is nevertheless *not* intentional under a discriminatory description. The implausibility of such a scenario indicates that the dominant feature differentiating intentional from unintentional discrimination is the satisfaction of the *second* condition. This is because in what seems to be a range of plausible cases, assuming that the basic conditions for a discriminatory action obtain (i.e., the action consists of a differential treatment underlain by a discriminatory mental state), it is enough that the second condition is satisfied for the action to constitute wrongful intentional discrimination, while the same is not true for the

---

<sup>35</sup> Gendler, 'Between Reason and Reflex'; Nanay, 'Between Perception and Action,' pp. 81-86.

first condition. The discussion of the moral difference between intentional and unintentional discrimination in the next section is based on this insight.

#### 5. The moral difference between intentional and unintentional discrimination

The conceptual difference between intentional and unintentional discrimination can be drawn, then, along the lines suggested above: the main feature separating the two is that in intentional discrimination the agent incorporates the content of her discriminatory belief into her motivating reason, that is, views it as a consideration counting in favor of her action, and guides her action based on this content in the way characteristic of acting based on motivating reasons. This is while in unintentional discrimination a discriminatory mental state only causally influences her action. Here, I argue that acting with the motivational structure characteristic of intentional discrimination involves endorsement of the underlying discriminatory belief; and that this particularly objectionable attitude is what generates the severity of intentional discriminatory actions.

This claim builds on some more basic assumptions about the role which discriminatory mental states or attitudes—not just intentions—play in generating the wrongness of both intentional and unintentional discriminatory actions. I have briefly illustrated the intuitive plausibility of the thought that discriminatory mental states are partly constitutive of the wrongness of such actions in Section 2; supporting the claim about the moral difference between the two types will benefit, however, from elaborating on its principled justification. This is discussed in subsection A, which also supports the general claim that whether an underlying objectionable attitude is intentionally acted on matters for the degree of wrongness of actions whose moral status is partly constituted by such attitudes. Subsection B then shows that whether the underlying objectionable attitude—particularly, a discriminatory belief—is intentionally acted on is significant for the moral appraisal of discriminatory actions in particular, as this plausibly involves endorsement of the belief's objectionable content.

##### A. Discriminatory mental states and the moral status of discriminatory actions

As the key examples of discriminatory actions mentioned throughout the discussion indicate, persistent intuitions maintain that the group-based differential treatment they involve is not enough to make them morally objectionable; rather, in both intentional and unintentional cases the action has to be accompanied or underlain by a discriminatory mental state for them to be objectionable in a characteristically discriminatory way. In that sense, discriminatory actions are similar to other actions performed in the course of interpersonal interactions, which seem objectionable in a way importantly tied to the agent's underlying mental states or attitudes—and not just their behavioral features (or their objectionable consequences; as we have seen, discriminatory actions often do not involve the latter).

A central example, often mentioned together with discriminatory actions to highlight this point, is that of publicly humiliating someone. Suppose, for instance, that I

share some potentially embarrassing information about a coworker in front of our mutual colleagues.<sup>36</sup> While in terms of behavior or consequences my action consists of just that—and can adequately be described as 'publicly sharing embarrassing information'—when it is accompanied by an attitude of condescension and indifference towards my colleague's feelings, the action arguably amounts to (objectionable) humiliation. Notice that for this to be the case I do not have to *intend* to embarrass my colleague, intentionally act on my condescending attitudes, or, more generally, have any malicious or otherwise objectionable intentions or motives; rather, it is enough that my action is accompanied or influenced by an objectionable attitude, to which my behavior of sharing the information is somehow appropriately related. Similar things can be said about, e.g., exercising my power as a governmental official to refuse someone's request in a bureaucratic process: while such refusal may be unobjectionable where it is within my position's authority, in a case where it is influenced by my attitude of megalomaniac desire for bureaucratic power and control, my action would constitute an objectionable abuse of power.

In such examples, the action's nature and its ethical features seem inextricably bound with the agent's underlying mental states or attitudes. In other words, what seems to matter for their moral assessment is not a 'thin' description of the agent's behavior, but a 'thick' description that also encompasses these underlying mental states or attitudes.<sup>37</sup> Thus, discriminatory actions (of the types discussed here) are appropriately characterized as such, that is, as involving an objectionable form of group-based differential treatment, in virtue of the agent's underlying discriminatory mental states or attitudes.

It is important to note that the negative moral appraisal of such actions does not have to amount to a determination that performing them is impermissible, or all-things-considered impermissible; in other words, the claim that the agent's underlying mental states transform an otherwise morally neutral or permissible action into an objectionable one does not entail that the action is rendered impermissible in virtue of these mental states. At a minimum, such a judgment does not seem appropriate where unintentional discriminatory actions are concerned—whereas those are plainly the appropriate target of some negative moral appraisal nevertheless, e.g., being morally criticizable, objectionable, wrongful, or bad. Similarly, performing intentional discriminatory actions might be permissible in some contexts despite being the appropriate target of such negative moral appraisal, e.g., when performing them is within the agent's legitimate scope of discretion. For example, it seems permissible for a homeowner to refuse to sell her house for any reason, despite the decision not to sell it to someone on account of her group membership being morally criticizable or bad.<sup>38</sup>

As some authors have commented, then, when actions of the type discussed here are concerned, i.e., those taking place in the course of interpersonal interactions, and whose

---

<sup>36</sup> The discussion and examples here draw from Hallvard Lillehammer, 'Scanlon on Intention and Permissibility,' *Analysis* 70 (2010), pp. 578-585.

<sup>37</sup> Lillehammer, 'Scanlon on Intention and Permissibility,' p. 580; Richard J. Arneson, 'What is Wrongful Discrimination,' *San Diego Law Review* 43 (2006), pp. 781-82; Laurence A. Blum, *Friendship, Altruism, and Morality* (New York: Routledge, 2009), pp. 99-104.

<sup>38</sup> Lillehammer, 'Scanlon on Intention and Permissibility,' p. 584; Steven Sverdlik, 'Motive and Rightness,' *Ethics* 106 (1996), pp. 341-44.



objectionable character is inextricably bound with the agent's underlying attitudes—as opposed, say, to actions involving very bad consequences or violation of the victim's strong, negative rights (e.g., killing of innocents)—a nuanced approach to their moral appraisal, which allows for judgments other than straightforward permissibility and impermissibility, seems called for. Discriminatory mental states seem significant for such moral evaluation, even if they do not play the particular role of rendering an otherwise permissible action impermissible.<sup>39</sup>

The thought that whether the agent *intentionally* acts on her objectionable attitudes matters in determining such actions' degree of severity (wrongness, badness, etc.) seems to flow naturally from the premise that objectionable attitudes are an important part of what generates these actions' wrongness to begin with. Thus, to continue the above discussion, if I intentionally act on my condescending attitudes and cause embarrassment to my colleague, that would constitute a particularly severe instance of humiliation; and if I act with a particular intention to subdue a person out of a megalomaniac sense of bureaucratic power and control, that would be a particularly objectionable instance of abuse of power. The next section shows that there are special reasons to think that this is the case where discriminatory actions are concerned, as intentionally acting on the underlying belief (by incorporating its content into one's motivating reasons) involves, as mentioned, a particularly objectionable attitude of endorsing its discriminatory content.

#### B. Endorsement of the underlying discriminatory belief and the particular severity of intentional discrimination

The notion of endorsement of mental states is commonly invoked in a variety of philosophical contexts, and so I assume that despite some possible controversy or unclarity at the margins, we are equipped with a sufficiently clear understanding of it. Roughly, endorsement consists of some form of positive evaluation or a pro-attitude directed at the mental state—that can be, *inter alia*, of epistemic or ethical nature and have emotional or affective components. In the case of belief, endorsement presumably involves a higher level of epistemic commitment or a more robust positive epistemic evaluation of its content relative to merely viewing it as truthful (the latter is, presumably, already present in simply holding the belief).<sup>40</sup>

---

<sup>39</sup> Patrick S. Shin, 'The Substantive Principle of Equal Treatment,' *Legal Theory* 15 (2009), pp. 170-71; Arneson, 'What is Wrongful Discrimination,' pp. 782-83. In that sense, the claim defended here is different from the one at issue in debates over the Doctrine of Double Effect (DDE). There, it is debated (very roughly) whether the agent's *intentions relative to a very bad outcome* her action brings about—the death of innocents—matters for the *permissibility* of her action. As the authors mentioned here have commented, this difference entails that debates over the DDE do not have straightforward implications for actions such as discrimination (cf. T.M Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge and London: Harvard University Press, 2008), pp. 58-60, 72-75, 88).

<sup>40</sup> Adam Leite, 'Changing One's Mind: Self-Conscious Belief and Rational Endorsement,' *Philosophy and Phenomenological Research* 97 (2018), pp. 150–171; Michael Huemer, 'Moore's Paradox and the Norm of Belief,' in Susana Nuccetelli and Gary Seay (eds.), *Themes From G. E. Moore: New Essays In Epistemology And Ethics* (Oxford: Clarendon Press, 2007), p. 148.

Why think that intentional discriminatory actions involve an attitude of endorsement towards the underlying discriminatory belief (or its content)? Before clarifying this point, it is important to note two things about the claim advanced here. First, I claim that acting in an intentionally discriminatory manner itself involves a tacit endorsement of the discriminatory belief; such endorsement does not necessitate conscious, deliberate reflection on the belief or its content (as discussed above, intentional discrimination does not require that the underlying belief is conscious or consciously entertained). Second, for our purposes it is sufficient to indicate that such an attitude is present at the point of performing a discriminatory action; I remain neutral on whether it is likely to be (or necessarily is, etc.) global and diachronically stable, that is, whether the agent would retain this attitude across a variety of activities and contexts, and preserve it through time (as mentioned in the introduction, that also means that I separate the moral appraisal of individual discriminatory actions from evaluations of the agent's character as a racist, sexist, etc.; the latter may be tied with such global and diachronically stable endorsement of discriminatory beliefs).<sup>41</sup>

The claim that in performing an intentionally discriminatory action an agent tacitly endorses her underlying discriminatory belief may be supported by examining the features of this belief and showing that there is a significant change in them—indicative of its endorsement—that is plausibly attributed to the performance of the action. Thus, it is useful to examine a case where an agent holds a discriminatory belief *not* endorsed prior to performing her action. Due to the socially-transmitted nature of many discriminatory beliefs, such cases are familiar and widespread. Consider, for instance, a slightly adjusted version of *Small Shop*. Suppose that the business owner (let us call her Rachel) has grown up in a society where anti-Roma sentiments and beliefs are widespread, but not commonly discussed in a direct and deliberate manner. Thus, it is plausible that Rachel has passively acquired many anti-Roma beliefs—for instance, associating the group with negative traits or prescribing offensive behaviors to it—due, *inter alia*, to her consistent perceptual exposure to stereotypical representations of the group, or participation in conversations in which inegalitarian propositions are vaguely or indirectly implied. Such beliefs are plausibly characterized as unendorsed.

Upon spotting a person of Roma descent attempting to enter her shop, Rachel promptly refuses to let her in, while shouting 'Roma out!' Plausibly, such a spontaneous action takes place without Rachel reflecting on her discriminatory belief and explicitly endorsing it. Nevertheless, the action seems to involve an important change in the belief's status, amounting to endorsement. Up to the point of performing the action, Rachel may have viewed the belief's discriminatory content as truthful; but in intentionally acting on it

---

<sup>41</sup> To clarify, then, this means that committed egalitarians, as well as committed racists or sexists (and anything in between) may perform intentional discriminatory actions, and vice versa: presumably, committed racists and sexists may *also* be implicitly biased, and when they act in a discriminatory manner influenced by these implicit biases their action is adequately characterized as *unintentional* discrimination. Relatedly, I do not assume that discriminatory beliefs endorsed in this manner should be associated with the agent's 'deep self' or any privileged part of her mental economy. Generally, the discussion is friendly to views on which agents' mental economy might be non-homogenous or 'fractured.' See Brownstein, *The Implicit Mind*, pp. 33-37.

she has moved to viewing this content as a consideration counting in favor of *acting*. That is, she adopts an attitude towards the discriminatory content that reflects a higher level of epistemic confidence or commitment relative to an unendorsed belief—she is willing to perform an intentional action that would be effective or rational only in a world where this content is truthful.

This point can be further supported by examining the nature of the practical commitment involved in performing such an action. I have suggested (in Section 3B) that due to their functional role, motivating reasons determine the nature of the 'subsidiary actions' of which the overall course of action is composed, and thus the content of the practical commitment embedded in a given intentional action (when it is characterized in a fine-grained manner); the latter includes a commitment to adjust the execution of one's intention in a way that appropriately matches the content of one's motivating reason(s). Such practical commitment involves, it seems, a particularly strong epistemic commitment to the belief whose content figures in one's motivating reason, akin to endorsing it. This is evident in actions that are somewhat temporally extended (such as a job interview): in performing such an action, the agent consistently adjusts her subsidiary actions to match the content of her underlying discriminatory belief, thus repetitively affirming her commitment to the belief's truthfulness and her willingness to view its content as a consideration counting in favor of acting.

Another way of supporting these claims is by showing that the features of beliefs tacitly endorsed in that manner resemble those of beliefs endorsed by conscious reflection. One important feature often associated with reflective endorsement is the agent's disposition to assert her reflectively endorsed beliefs. A similar feature is present when one incorporates the content of her belief in her motivating reason. As noted above, an agent is disposed to present the content of her motivating reason in response to the question, 'Why did you act the way you did?' A natural way of understanding an answer to such a question is not as a mere presentation of the belief's content; rather, it should be understood as *asserting* it. In saying, for example, 'I acted the way I did because Roma should not be associated with,' I do not only present the proposition that guided me in executing my intention in a neutral way; I am asserting that 'Roma should not be associated with!' in a way indicating endorsement.

As can be seen, then, in acting with the motivational structure characteristic of intentional discrimination the agent exhibits a particularly objectionable discriminatory attitude, that is, she endorses her discriminatory belief, or has a high degree of epistemic commitment to its truthfulness. This attitude is markedly more objectionable relative to the discriminatory attitude underlying unintentional discriminatory actions, where at most a discriminatory belief only causally influences the agent's behavior, and, considering that implicit biases may consist in something other than beliefs (e.g. conceptual associations) the agent might not even believe a discriminatory proposition. Considering that the underlying discriminatory attitude is what generates the moral objection to discriminatory actions (in conjunction with the group-based differential treatment), it is plausible that this generates the particular severity of intentional discriminatory actions, making them significantly more objectionable relative to unintentional ones.

## 6. Conclusion

Despite the prominence of the distinction between intentional and unintentional discrimination in the literature and public discourse, basic conceptual and ethical questions surrounding it remain obscure. The paper has attempted to take the first step towards a clearer, more philosophically-informed discussion of the topic. It has argued that the ethical difference between the two types is traceable to a difference in the agent's underlying attitude: while in intentional discrimination discriminators endorse a discriminatory proposition (by seeing it as a consideration that counts in favor of their action), in unintentional discriminatory actions they are only causally influenced by a discriminatory mental state—making them act in a manner consistent with a discriminatory proposition without endorsing it. Recent developments in the literature and the political landscape have reinvigorated debates to which the distinction is relevant; the discussion here will, I hope, contribute to gaining a better understanding of the ethical stakes involved in these debates.

## **Acknowledgements**

This paper is a revised version of part of my dissertation project. I am grateful to my PhD supervisors David Enoch and Moshe Halbertal for their comments and suggestions, and for their general help and support throughout the process of working on prior versions. I also thank Andrew Altman and Kasper Lippert-Rasmussen for written comments, Hanoch Dagan and participants of the Edmond J. Safra Center for Ethics Colloquium (Tel Aviv University) for helpful comments and discussion, and an anonymous reviewer for this Journal, whose comments and suggestions greatly improved the paper

## **Biographical Note**

Rona Dinur is a Visiting Max Weber Fellow at the European University Institute. She holds a PhD in philosophy and an LLB/BA in law and political science from the Hebrew University of Jerusalem, and an LLM from Harvard Law School.

### References

- Adams, Frederick (1986) Intention and Intentional Action: The Simple View. *Mind & Language* 1.4: 281–301.
- Alexander, Larry (1992) What Makes Wrongful Discrimination Wrong? Biases, Preferences, Stereotypes, and Proxies. *University of Pennsylvania Law Review* 141.1: 149–219.
- Altman, Andrew (2016) Discrimination. In Edward N. Zalta ed., *The Stanford Encyclopedia of Philosophy*. Available at:  
<<http://plato.stanford.edu/archives/win2016/entries/discrimination>>
- Alvarez, María (2009) Acting Intentionally and Acting for a Reason. *Inquiry* 52.3: 293–305.
- (2010) *Kinds of Reasons: An Essay in The Philosophy Of Action*. Oxford.
- (2017) Reasons for Action: Justification, Motivation, Explanation. In Edward N. Zalta ed., *The Stanford Encyclopedia of Philosophy*. Available at:  
<https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/>.
- Anscombe, G. E. M. (1957) *Intention*. 2nd Ed, Cornell University Press.
- Arneson, Richard J. (2006) What is Wrongful Discrimination. *San Diego Law Review* 43.4: 775–808.
- Audi, Robert (1991) Intention, Cognitive Commitment, and Planning. *Synthese* 86.3: 361–378.
- Bagnis, Arianna, Alessia Celeghin, Cristina Onesta Mosso, and Marco Tamietto (2019) Toward an Integrative Science of Social Vision in Intergroup Bias. *Neuroscience and Biobehavioral Reviews* 102: 318–326.
- Beeghly, Erin (2015) What is a stereotype? What is stereotyping? *Hypatia* 30.4: 675–691.
- Blum, Lawrence (2004) Stereotypes and Stereotyping: A Moral Analysis. *Philosophical Papers* 33.3: 251–289.
- Blum, Laurence A (2009) *Friendship, Altruism, and Morality*. Routledge.
- Bratman, Michael (1987) *Intention, Plans, and Practical Reason*. Harvard University Press.
- Brownstein, Michael (2018) *The Implicit Mind: Cognitive Architecture, the Self, and Ethics*. Oxford University Press.
- Brownstein, Michael and Jennifer Saul (2016) Introduction. In Michael Brownstein and Jennifer Saul, eds. *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*. Oxford University Press, 1–19.
- Craig, Ronald (2007) *Systemic Discrimination in Employment and the Promotion of Ethnic Equality*. Martinus Nijhoff Publishers.
- Davison, Donald (1963) Actions, Reasons, and Causes. *The Journal of Philosophy* 60.23: 685–700.

- Dovidio, John F., Kerry Kawakami, and Samuel L. Gaertner (2002) Implicit and Explicit Prejudice and Interracial Interaction. *Journal of Personality and Social Psychology* 82.1: 62–68.
- Eidelson, Benjamin (2015) *Discrimination and Disrespect*. Oxford University Press
- Fazio, Russell H., Joni R. Jackson, Bridget C. Dunton, and Carol J. Williams (1995) Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: a Bona Fide Pipeline? *Journal of Personality and Social Psychology* 69.6: 1013–1027.
- Gendler, Tamar Szabó (2011) On the Epistemic Costs of Implicit Bias. *Philosophical Studies* 156.1: 33–63.
- (2012) Between Reason and Reflex: Response to Commentators. *Analysis* 72.4: 799–811.
- Greenwald, Anthony G., and Mahzarin R. Banaji (1995) Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes. *Psychological Review* 102.1: 4-27.
- Hellman, Deborah (2008) *When Is Discrimination Wrong?* Harvard University Press.
- Huemer, Michael (2007) Moore's Paradox and the Norm of Belief. In Susana Nuccetelli and Gary Seay eds., *Themes From G. E. Moore: New Essays In Epistemology And Ethics*. Clarendon Press, 142-157.
- Huq, Aziz Z. (2018) What Is Discriminatory Intent? *Cornell Law Review* 103.5: 1211–1292.
- Leite, Adam (2018) Changing One's Mind: Self-Conscious Belief and Rational Endorsement. *Philosophy and Phenomenological Research* 97.1: 150-71.
- Levy, Neil (2015) Neither Fish nor Fowl: Implicit Attitudes as Patchy Endorsements. *Noûs* 49.4: 800–823.
- Lillehammer, Hallvard (2010) Scanlon on Intention and Permissibility. *Analysis* 70.3: 578-585.
- Lippert-Rasmussen, Kasper (2012) Discrimination and Equality. In Andrei Marmor ed., *The Routledge Companion to Philosophy of Law*. Routledge, 569-83.
- (2013) *Born Free and Equal? A Philosophical Inquiry into the Nature of Discrimination*. Oxford University Press.
- Locke, Dustin (2015) Knowledge, Explanation, and Motivating Reasons. *American Philosophical Quarterly* 52.3: 215–232.
- Madva, Alex and Michael Brownstein (2018) Stereotypes, Prejudice, and the Taxonomy of the Implicit Social Mind. *Noûs* 52.3: 611–644.
- Mandelbaum, Eric (2016) Attitude, Inference, Association: On the Propositional Structure of Implicit Bias. *Noûs* 50.3: 629–658.
- McCann, Hugh J. (1991) Settled Objectives and Rational Constraints. *American Philosophical Quarterly* 28.1: 25-36.
- Mele, Alfred R. (2003) *Motivation and Agency*. Oxford University Press.

- Nanay, Bence (2013) *Between Perception and Action*. Oxford University Press.
- Payne, B. Keith (2001) Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon. *Journal of Personality and Social Psychology* 81.2: 181–192.
- Payne, B. Keith, and Heidi A. Vuletich (2018) Policy Insights from Advances in Implicit Bias Research. *Policy Insights from the Behavioral and Brain Sciences* 5.1: 49-56.
- Pollard, Bill (2006) Explaining Actions with Habits. *American Philosophical Quarterly* 43.1: 57-69.
- Scanlon, T. M. (2008) *Moral Dimensions: Permissibility, Meaning, Blame*. Belknap Press of Harvard University Press.
- Schwitzgebel, Eric (2010) Acting Contrary to Our Professed Beliefs or the Gulf Between Occurrent Judgment and Dispositional Belief. *Pacific Philosophical Quarterly* 91.4: 531–553.
- Searle, John R. (1983) *Intentionality: an Essay in the Philosophy of Mind*. Cambridge University Press.
- Setiya, Kieran (2011) Knowledge of Intention. In Anton Ford, Jennifer Hornsby, and Frederick Stoutland eds. *Essays on Anscombe's Intention*. Harvard University Press, 170–197.
- Shin, Patrick S (2009) The Substantive Principle of Equal Treatment. *Legal Theory* 15.2: 149-72.
- Sverdlik, Steven (1996) Motive and Rightness. *Ethics* 106. 2: 327-349.
- Tuomela, Raimo (2007) Motivating Reasons for Action. In Mark Timmons, John Greco and Alfred R. Mele eds. *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi*. Oxford University Press, 176-98.
- Valaris, Markos (2015) The Instrumental Structure of Actions. *The Philosophical Quarterly* 65.258: 64-83.
- Velleman, James David (200) *The Possibility of Practical Reason*. Clarendon.
- Washington v. Davis*, 426 U.S. 229 (1976)