Journal of the American Philosophical Association (2024) 295–312 © The Author(s), 2023. Published by Cambridge University Press on behalf of the American Philosophical Association
DOI:10.1017/apa.2023.5

Biased Evaluative Descriptions

ABSTRACT: *In this essay I identify a type of linguistic phenomenon new to feminist philosophy of language: biased evaluative descriptions. Biased evaluative descriptions are descriptions whose well-intended positive surface meanings are inflected with implicitly biased content. Biased evaluative descriptions are characterized by three main features: (1) they have roots in implicit bias or benevolent sexism, (2) their application is counterfactually unstable across dominant and subordinate social groups, and (3) they encode stereotypes. After giving several different kinds of examples of biased evaluative descriptions, I distinguish them from similar linguistic concepts, including backhanded compliments, slurs, insults, epithets, pejoratives, and dog whistles. I suggest that the traditional framework of Gricean implicature cannot account for biased evaluative descriptions. I discuss some challenges to the distinctiveness and evaluability of biased evaluative descriptions, including intersectional social identities. I conclude by discussing their social significance and moral status. Identifying biased evaluative descriptions is important for a variety of social contexts, from the very general and broad (political speeches) to the very particular and small (bias in academic hiring).*

In 2008, Joseph Biden called Barack Obama 'an African-American who is articulate, bright and clean and a nice-looking guy' (NPR 2007). Articulateness was often attributed to four-star African American military general and Secretary of State Colin Powell. In the present day, former democratic presidential candidate Pete Buttigieg, an openly gay married man, has often been complimented in the press on how 'traditional' he is said to be (Smith 2019). Trans actress Laverne Cox is overly described as 'gorgeous' by well-meaning fans (Allen 2017). Though *articulate*, *clean*, *traditional*, and *gorgeous* are intended as compliments, arguably they would not be applied in similar situations to people belonging to different, more dominant social categories—that is, if Barack Obama were not Black, if Pete Buttigieg were not gay, and if Laverne Cox were not trans. Such *biased evaluative descriptions*—roughly, well-intended descriptions whose apparently positive surface meanings are inflected with implicit bias or benevolent discrimination—are the focus of this essay.

Thanks to Louise Antony, Ray Briggs, Bianca Cepollaro, Esa Díaz-León, Anne Eaton, Faeze Faezeli, Carolina Flores, Dan López de Sa, Daniel Nolan, Dee Payton, Michael Rea, Jennifer Saul, Adam Sennet, Georgette Sinkler, Devin Sprague Morse, Isidora Stojanovic, and several anonymous referees for excellent feedback on this essay. Thanks also to audiences at the Words Workshop, the Pacific APA, the University of Illinois- Chicago, the University of Barcelona, Notre Dame Gender Studies, Rutgers University, the University of St Andrews, and the Institut Jean Nicod for excellent questions. I owe gratitude to my feminist philosophy class at the University of Notre Dame for hashing out these ideas with me. Thanks also to the National Endowment for the Humanities: revisions on this essay were completed during my fellowship year.

My primary goals of the discussion are to characterize the phenomenon and to make headway in diagnosing the linguistic force and moral significance of such descriptions. My secondary goal is to draw attention to a heretofore overlooked topic in feminist philosophy of language: well-intentioned discriminatory speech. Feminist philosophy of language has a long tradition of analyzing straightforwardly negative speech, like slurs and epithets. Feminist speech act theory has been extensively utilized in exploring the illocutionary force and moral status of pornography. But philosophers of language and feminist philosophers have not turned their attention to the pervasive phenomenon of well-intentioned speech inflected with sexism and other discriminatory attitudes. Here I aim to lead the way.

1. Demarcating the Phenomenon

Biased evaluative descriptions are a species of a broader genus of linguistic phenomena infused with implicit bias. There are questions whose occurrences betray bias, such as when journalists ask women athletes about their personal lives. There are biased linguistic omissions, as when a letter of reference discusses a woman's personality while omitting her accomplishments. There are biased appraisals, as when women are verbally scrutinized for signs of competence more than men. And there are likely many more similar sorts of biased evaluative phenomena. Though my focus on biased evaluative descriptions is relatively narrow, some of my remarks apply to these other species of biased linguistic phenomena.

That the focus is on biased evaluative *descriptions* does not imply that bias is somehow intrinsic to the descriptions themselves. There is nothing about specific lexical items like the adjective *articulate* that makes them particularly likely to be infused with bias. Rather, there is a pattern of use of such descriptions that reflects implicit biases, and it is this pattern of use in which I am interested. (For a discussion of the broader related phenomenon of linguistic 'microaffirmations', see Delston [2021].)

Call the examples with which I began *positive* biased evaluative descriptions. Positive biased evaluative descriptions are intended to be complimentary. They are to be contrasted with negative biased evaluative descriptions, such as *shrill* (said of Hillary Clinton) and *flamboyant* (said of many gay men)—descriptions that are products of openly hostile sexism and racism. Such negative descriptions would not be applied if their targets were not members of certain social categories, even holding fixed other relevant attributes of their targets. But because I am interested in well-intentioned sexist speech rather than speech backed by overt sexism and discrimination, here I focus primarily on positive biased evaluative descriptions.

Biased evaluative descriptions encompass a broad but overlapping range of phenomena. Some are primarily *counterstereotypical*: their use is intended to negate stereotypes associated with particular social groups. The examples with which I began are counterstereotypical biased evaluative descriptions. Biden intended to counter a stereotype of African American politicians as somehow inarticulate, unclear, or linguistically incompetent. In calling Pete Buttigieg *traditional*, the well-meaning press presumably intends to negate a stereotype of gay men as promiscuous and transgressive. And in calling Laverne Cox *gorgeous*, fans intend to counter the stereotype of trans women as unattractive *qua* women. An overlapping phenomenon is the biased compliment, such as 'Yasss king!' uttered by cis persons towards trans men as well-intentioned affirmation of the latter's masculinity. A hijab-wearing student of mine laughed about how frequently she is complimented on being *open-minded*: 'I *am* open-minded', she explained, 'but very few people would go out of their way to say this about me if I did not wear the hijab'.[1]

Other biased evaluative descriptions are *counternormative*. Users of counternormative biased evaluative descriptions intend to counteract pernicious norms by directly opposing them. A ubiquitous case is the propensity of white women to indiscriminately call Black women beautiful, in an apparent attempt to oppose narrow norms of Caucasian beauty. Though I have witnessed this phenomenon many times in real life, I was also pleased to see it reflected in a fictional exchange between Nigerian-American character Ifemelu and her white acquaintance Kimberly in Chimamanda Adichie's autobiographical fiction *Americanah* (2013):

> Ifemelu would come to realize later that Kimberly used 'beautiful' in a peculiar way. 'I'm meeting my beautiful friend from graduate school', Kimberly would say, or 'We're working with this beautiful woman on the inner-city project', and always, the women she referred to would turn out to be quite ordinary-looking, but always black. One day, late that winter, when she was with Kimberly at the huge kitchen table, drinking tea and waiting for the children to be brought back from an outing with their grandmother, Kimberly said, 'Oh, look at this beautiful woman', and pointed at a plain model in a magazine whose only distinguishing feature was her very dark skin.
>
> 'Isn't she just stunning?'
>
> 'No, she isn't'. Ifemelu paused. 'You know, you can just say 'black'. Not every black person is beautiful'.
>
> Kimberly was taken aback, something wordless spread on her face and then she smiled, and Ifemelu would think of it as the moment they became, truly, friends. (2013: 180–81)

In this context, *beautiful* functions as a counternormative biased evaluative description. The term would not be used if the woman were not Black, and it is used in service to dispelling the myth that Black women are not beautiful. Kimberly, the user of the term, is well intentioned. But such a compliment has its roots in a form of benevolent racism (which I discuss below) according to which all women belonging to a certain racial category count as beautiful—an essentializing, oversimplified claim.

[1] Thanks to Fafa Faezeli for this example and for allowing me to use it.

Other biased evaluative descriptions are primarily *diminutive*: even while they are intended to be straightforwardly complimentary, their positive force is diminished due to the comparison class of descriptions used for members of non-marginalized social categories. For example, the adjective most often used to compliment my professional philosophy talks is *fun*. My talks *do* tend to be fun: they contain interesting examples and snappy jokes, and they are constructed with clarity and accessibility in mind. I am not at all offended by the *fun* label; in fact, I pride myself on giving lively, engaging talks. But *fun* can diminish the other, more professionally valuable aspects of a talk (clarity, explanatory power, creativity, intellectual depth) in favor of an aspect considered less important to professional prestige. Further, the frequency with which the label is applied to my talks contrasts with the apparent sparsity with which it is applied to similar talks given by male colleagues.

Diminutive biased evaluative descriptions are pervasive in letters of recommendation for women. A prominent study by Frances Trix and Carolyn Psenka (2003) on differences between letters of recommendation for male and female medical faculty candidates found a significant increase in what they term 'grindstone adjectives' (2003: 207), or adjectives used to describe being hardworking, in letters for women. Along with *hardworking*, typical examples of grindstone adjectives include *conscientious* and *diligent* (Trix and Psenka 2003: 207). Being hardworking is obviously a good trait for any faculty member to possess, but the description is primarily used as a contrast class for the apparent possession of natural talent and innate genius. Judgments of natural talent are often deeply inflected with racial and gender bias. For example, Meredith Meyer, Andrei Cimpian, and Sarah-Jane Leslie (2015) show that fields in which natural talent is thought to play a role are overwhelmingly dominated by white men, and such judgments notoriously track social and physical traits of this population. Even apparently straightforward descriptions such as *accomplished* and *professional* can count as diminutive biased evaluative descriptions, depending on context and utterer.

Many biased evaluative descriptions span more than one category: some are both counternormative and counterstereotypical, and some are counterstereotypical and diminutive. As applied to Barack Obama, *articulate* is counterstereotypical and diminutive. As applied to Pete Buttigieg, *traditional* is counternormative and counterstereotypical. Another common sort of category-spanning example is the propensity of men heavily engaged in child-rearing to be commended on how *involved* they are said to be. Intended as a compliment, the description is meant to counteract the stereotype of men as distant fathers. But the description is also diminutive insofar as it stands in contrast to stronger, unqualified compliments and evaluations of their parenting activities and abilities. That biased evaluative descriptions can span categories does not diminish the explanatory power of the categories themselves, since the categories help us understand biased evaluative descriptions and more carefully identify their effects.

Finally, while my focus is biased evaluative descriptions that are applied to marginalized groups, there are numerous descriptive utterances that fit the bill even when directed at people in dominant social groups. Consider one friend saying to another of a party invitee: ‘He’s white, but don’t worry, he’s cool’.[2] Here, the use of “white” is counterfactually unstable across dominant and subordinate social groups, but the role of the groups is switched from the other canonical examples. Similarly with ‘She’s rich, but she’s not stuck up’. These are interesting in their own right, but they are not the topic of my attention.

As suggested by these examples and the different sorts of categories under which they fall, most biased evaluative descriptions have three major elements: (1) they involve implicit bias or benevolent discriminatory attitudes, (2) they are counterfactually unstable across subordinate and dominant social identities, and (3) they involve or encode stereotypes.

1.1 Implicit Bias and Benevolent Discriminatory Attitudes

Many positive biased evaluative descriptions are products of implicit bias. *Implicit bias* encompasses a set of unconscious or subconscious attitudes, beliefs, and stereotypes that influence thought and behavior. Implicit bias is ubiquitous and near universal, and spans political, religious, and philosophical belief systems. Men and women suffer from implicit bias that targets women; people of all races suffer from implicit biases targeting non-white people. (See Jost et al. [2009] for concrete examples of the ubiquitous harm of implicit bias.)

Crucially, implicit biases often do not line up with explicitly endorsed beliefs: most users of positive biased evaluative descriptions would not explicitly endorse racist or sexist principles, and would be surprised to learn that their use of biased evaluative descriptions signals such biases. This point is important for zeroing in on the sort of apparently positive descriptions that are the subject of my investigation.

Peter Glick and Susan Fiske (1996) draw a distinction between *hostile sexism* and *benevolent sexism*. Hostile sexism targeted at women is undergirded by explicitly negative and reductionist attitudes toward women—think Rush Limbaugh, Ann Coulter, and *The Handmaid’s Tale*. Benevolent sexism, in contrast, is a set of sexist attitudes that masquerades as friendly pro-woman ideology. Benevolent sexism is a form of bias and most often a form of implicit bias. (Here I associate hostile sexism with explicit bias and benevolent sexism with implicit bias, but this is not always the case: explicit bias can be benevolent, and implicit bias is often malevolent. Thanks to a referee for pointing this out.) According to Glick and Fiske, ‘Benevolent sexism is a set of interrelated attitudes toward women that are sexist in descriptions of viewing women stereotypically and in restricted roles but that are subjectively positive in feeling tone (for the perceiver) and also tend to elicit behaviors typically categorized as prosocial (e.g., helping) or intimacy-seeking (e.g., self-disclosure)’ (1996: 491). In another article, Glick and colleagues state that benevolent sexism is ‘a subjectively positive orientation of protection, idealization, and affection directed toward women that, like hostile sexism, serves to justify women’s subordinate status to men’ (Glick et al. 2000: 763).

In other words, many instantiations of benevolent sexism that are intended to be positive or supportive of women in fact reinforce the subordinate social status of

[2] Thanks to a member of the audience at the Rutgers Feminist Philosophy Reading Group for this example.

women. Benevolent sexist attitudes include paternalistic attitudes, among them assumptions of women's emotional fragility (for example, women must be emotionally protected because they are more pure), essentialist views about women's supposed goodness (women are naturally more compassionate and nurturing), and reductive views about women's dispositions and abilities (women are less aggressive than men, which makes them less suited to aggressive questioning in such settings as academic philosophy.) A significant amount of positive political coverage of Elizabeth Warren in the 2020 presidential race implied that a woman would automatically be a better president than a man because women are naturally more compassionate and reasonable.

Benevolent sexism is a progenitor of many well-intentioned positive biased evaluative descriptions, and is pervasive among well-meaning self-identified allies of women and minorities. According to a study by Ivona Hideg and D. Lance Ferris (2016), holders of benevolent sexist attitudes are more likely than others to support equal opportunity policies in the workplace. Users of positive biased evaluative descriptions intend to be supportive and well-meaning, and often work toward genuine social good.

There are also numerous examples of benevolent racism, benevolent neurotypicality, and other forms of benevolent bias. Asian-American workers in Silicon Valley are underrepresented at the upper echelons of various companies because they are perceived to have already made it compared to other racial minorities (Schiavenza 2018). People on the autism spectrum complain of being stereotyped as mathematical or scientific savants (McGrath 2019). Jewish lawyers complain that they are hired because they are thought to be more effective at practicing law than others (Jones 2010).

The manner and extent to which a biased evaluative description is caused by implicit bias varies by type. Counterstereotypical and counternormative biased evaluative descriptions often involve a conscious effort to counteract stereotypes and norms, as with Kimberly in the fictional case from Adichie's *Americanah*. Diminutive biased evaluative descriptions often involve unencumbered bias that is buried much deeper. In the case of a woman's talk labeled *fun*, for example, the user's choice of words can stem from implicit bias of which they are not aware. That is, they associate good talks by women with being fun, while associating good talks by men with clarity and intellectual depth. An entire article could be written about these variations alone, and I cannot do justice to the topic here. I hope that this essay is the beginning of the investigation into these phenomena.

1.2 Counterfactual Instability across Subordinate and Dominant Social Groups

Many positive biased evaluative descriptions express an evaluation of a person that would not be applied were a particular sort of subordinate social identity involving gender, race, ability, socioeconomic status (etc.) not occupied by the person being evaluated. I call this property of biased evaluative descriptions *counterfactual instability across social identities*. A helpful way of assessing whether an evaluative description is biased is to swap in a socially dominant identity for a socially subordinate identity, while holding other relevant things fixed. For example, Barack Obama would not have been complimented on his articulateness were he not a member of a particular racial minority: the same description would not be applied to an identically skilled white candidate (or at least, not as often). Entertaining such *countersocial counterfactuals*, counterfactuals that run contrary to social fact, is an intuitive method of identifying biased evaluative descriptions. (See Bernstein [manuscript] for more extensive discussion of countersocials.)

While we often seem to entertain countersocial counterfactuals fairly easily ('Would he have said that if I were a man?'), the reasoning behind such evaluations is surprisingly complicated. Countersocial variations are complex because subordinate social categories are not just labels. Subordinate social categories are backed by stereotypes and conceptually rich ideologies. In counterfactually evaluating the Obama example, it is not enough to imagine that everything is the same but that Barack Obama just has less melanin. We would be holding fixed too much. That simple counterfactual evaluation ignores the details of social constructions of race, and associated ideologies. The possible world that we entertain is the one that varies Barack Obama's name, mode of presentation, and arguably his political roots on the South Side of Chicago. We somehow imagine that he inhabits not just a different physical form, but that he inhabits a social category with very different extrinsic features and relationships to American society. Similarly for other such countersocial evaluations, which are not just a matter of swapping out physical traits, colors, or parts, as if one is playing with a Mr. Potato Head.[3]

Whether or not something is a biased evaluative description is not a matter of the evaluation's truth or falsity in a given instance. It is true, for example, that Barack Obama is articulate. Many of my talks *are* actually fun. Pete Buttigieg *is* fairly traditional. Women described as *hardworking* in letters of recommendation presumably *are* hardworking.

'What's wrong with saying she's hardworking?' is a common refrain among those who use this and other grindstone adjectives. What is wrong is that *hardworking*, *energetic*, and many similar positive biased evaluative descriptions stand in contrast to descriptions of greater professional value applied to more socially dominant groups. Grindstone adjectives are biased because they, rather than other descriptions, are applied to members of socially subordinate categories while other, conventionally stronger descriptions are applied to members of socially dominant categories. The problem is not with the descriptions themselves, but with their differing patterns of use across dominant and subordinate social groups. A contrastive structure is especially common among diminutive biased evaluative descriptions: if a member of a dominant group with the same qualifications would be described by certain kinds of descriptions (for example, those involving natural talent like *gifted*) rather than other sorts of descriptions indicating traits of less professional value (for example, those alluding to hard work like *persistent*), then the latter descriptions are likely to exhibit bias, *ceteris paribus*. Obviously, professional values vary by context. *Hardworking* and

[3] It is hard to give a good story about how we manage to evaluate countersocial counterfactuals so reliably. In my discussion, I assume that we can broadly agree on how to evaluate them.

persistent in the mining industry will have different implicatures than *hardworking* and *persistent* in academic philosophy.

1.3 Encoding of Stereotypes

Finally, most biased evaluative descriptions encode or involve stereotypes. Some do so fairly explicitly. Hearing Buttigieg described as 'traditional' immediately calls up mental images of the contrast class of nontraditional gay men—tight clothes, wild parties in the Castro District, and so on. In context, *traditional* implies that 'Pete Buttigieg is traditional (for a gay man)' or 'Pete Buttigieg is traditional (whereas many gay men are not)'.

Other biased evaluative descriptions encode stereotypes more covertly. In the context of a letter of recommendation, a *hardworking* woman calls to mind someone who seeks to overcome her lack of natural talent with sheer grit. *Fun*, as applied to an academic talk, calls to mind a contrast class of a stereotypical philosopher (usually older, white, and bearded) droning on about a technical topic from the podium without looking up from his notes. The description at once grants faint praise on the target, while essentializing and stereotyping extant members of the target's social category. One thereby reinforces the pernicious norms and expectations that one intends to dissolve with the purportedly positive evaluation.

1.4 What is Problematic about Biased Evaluative Descriptions?

The above discussion helps get a handle on the phenomenon of biased evaluative descriptions, but it does not entirely capture what is harmful about them. I have alluded to the idea that they are problematic because they figure into different patterns of application between dominant and marginalized social groups. But this is not the entire explanation, since mere differences in patterns of use are not always bad. Children are described differently than adults, for example, and papers by students are described differently than papers by colleagues.

Biased evaluative descriptions are harmful in several ways. First, biased evaluative descriptions can create or reinforce low expectations for their particular subjects and for fellow members of marginalized groups. For example, if 'articulate' is considered the highest form of compliment for a Black candidate but not for a non-Black candidate, this differential usage plays a role in shaping perceptions of a particular Black candidate's potential, and of the potential of Black political candidates more generally. Biased evaluative descriptions treat as surprising or remarkable that a member of a minoritized group has a particular capacity or trait—one that would not be considered remarkable for members in socially dominant groups. To the extent that language shapes concepts and expectations of social groups, biased evaluative descriptions thus play a role in negatively shaping concepts of already-marginalized social groups.

Second, biased evaluative descriptions can lead to harms subsequent to their use. (See Bingeman [manuscript] for discussion of the moral risks of these forms of praise.) Members of marginalized groups are bound to miss out on jobs and opportunities when they are caged by comparatively low expectations: for better or for worse, academic positions are more likely to go to candidates described as *brilliant* than to those described as *hardworking*. If a specific sort of term is primarily applied to members of socially dominant groups in professional contexts, there is a tendency to judge all candidates based on their similarity to the socially dominant group.

Biased evaluative descriptions can also do harm through their encoding of stereotypes. As I suggest above, some biased evaluative descriptions pit their subjects against fellow members of marginalized social groups. For example, calling a hijab-wearing Muslim woman *open-minded* evokes a contrastive stereotype of dogmatic hijab-wearing women. On the other hand, indiscriminate application of terms, such as *beautiful* for all Black women, deprives the entire social group of more nuanced judgments.

Invoking stereotypes by directly combating them also risks subjecting targets to Marilyn Frye's famous 'double bind of oppression' (1983: 1–16), roughly a situation in which there is no right way to act as a member of a marginalized group. For example, women political candidates who are seen to possess stereotypically feminine traits such as warmth are thus considered too weak to do the job, but women candidates who are not perceived as warm are seen as too aggressive for political deal making. Democratic commentators complained that Pete Buttigieg was too traditional to be considered a genuinely queer political candidate: he was too gay for the Right, and too traditional for the Left (Downs 2019). Stereotypes socially punish their subjects whether or not the subjects conform to expectations.

It is not surprising that certain evaluative descriptions can do harm. But it is surprising that these harms can be created through utterances that are intended to be compliments. As I now discuss, one reason that these harms have been underexplored is that biased evaluative descriptions do not easily fit into existing and heavily studied linguistic taxonomies.

2. Biased Evaluative Descriptions and Existing Linguistic Frameworks

Biased evaluative descriptions are to be distinguished from several nearby linguistic phenomena, including straightforward compliments, backhanded compliments, euphemisms, slurs, insults, epithets, pejoratives, and dog whistles. Though biased evaluative descriptions share similarities with many of these things, they are also importantly different.

Biased evaluative descriptions are not straightforward compliments. Straightforward compliments are uncomplicatedly positive, such as 'He is the most talented politician I have ever encountered' or 'He is a great politician'. Straightforward compliments are neither stereotype invoking nor counterfactually unstable across dominant and subordinate social groups. They do not have contextually salient negative contrast classes.

Distinguishing biased evaluative descriptions from backhanded compliments is a trickier matter. Generally, backhanded compliments are intended to be cutting or slightly insulting, as in 'You look good for your age' or 'You're a good weight-lifter for a woman'. Backhanded compliments are sometimes counterfactually unstable across dominant and subordinate social groups. But backhanded compliments can be distinguished from biased evaluative descriptions via speaker intention: positive biased evaluative descriptions are intended to be positive, whereas backhanded compliments are not. Biased evaluative descriptions might *function* as backhanded compliments if an audience is disposed to read them as such, as in the case of letters of recommendation. But canonical backhanded compliments are intended to be cutting or insulting; biased evaluative descriptions are not.

Biased evaluative descriptions are not slurs or epithets. Both slurs and epithets encode intentionally negative content, explicitly target and essentialize members of social groups, and are intended to be pejorative or offensive. Labeling a woman a *slut* in a non-reclamatory context, for example, is intended as a mode of sexualized shaming; similarly for the use of *fag* for a gay person. In using a slur in a non-reclamatory context, a person endorses its offensive content (see, for example, Bolinger [2017] for a pragmatic account of slurs, and see Popa-Wyatt and Wyatt [2018] for an account of slurs that incorporates dominant and submissive social roles). For similar reasons, biased evaluative descriptions are not pejoratives, which convey intentionally negative content. (For recent accounts of pejoratives, see Hom [2010]; Sennet and Copp [2015]; and Marques and García-Carpintero [2014].) Utterers of positive biased evaluative descriptions do not explicitly endorse negative or pejorative content.

Biased evaluative descriptions are also to be distinguished from dog whistles, which are specifically designed to encode derogatory content for a private audience who understands the code, as when contemporary politicians call Jewish persons *cosmopolitan*. Users of positive biased evaluative descriptions generally do not intend to encode negative or stereotypical content, and the audience for such descriptions is public.

Positive biased evaluative descriptions are not straightforward insults, which intentionally communicate negative information or lack of respect about the target. It might turn out that some biased evaluative descriptions are unintentional or non-straightforward insults, if there are such things. (See Daly [2018] for a view of insults as expressions of lack of due regard.) Well-intentioned users of biased evaluative descriptions intend to communicate positive features of the target.

Biased evaluative descriptions are not euphemisms, which indirectly name a trait or a cluster of traits. For example, *electability* in the 2020 American democratic primary election encoded male traits. Racist media directed toward Meghan Markle that labels her *exotic* encodes African American traits. But biased evaluative descriptions straightforwardly attribute traits rather than shrouding them in euphemisms.

Some biased evaluative descriptions have features in common with scalar implicature. Roughly, there is scalar implicature when an utterer's choice of an informationally weak term over an informationally stronger term communicates something beyond surface meaning (Rett [2020]; Schlenker [2012]; Hirschberg [1985]). For example, calling an aspiring graduate student "punctual" in a letter of recommendation implies that the student does not have stronger academic traits than punctuality. Some diminutive biased evaluative descriptions function through scalar implicature insofar as the choice of one term over another communicates something in addition to the surface meaning of the term. In cases of diminutive biased evaluative descriptions, choosing one sort of term over another for a marginalized group is closely related to why the use of such terms is problematic. But while scalar implicature can help illuminate the mechanisms behind some biased evaluative descriptions, scalar implicature does not perfectly align with the phenomenon, for two reasons. First, not all biased evaluative descriptions fall on a single informational spectrum. For example, *articulate* is not necessarily a lower-information term than *brilliant*. Second, many canonical examples of scalar implicature involve deliberate communication of extra content, in contrast to the non-deliberate character of our target phenomenon.

It might be tempting to try to explain biased evaluative descriptions in terms of Grice's (1989) theory of implicature, which famously distinguishes between what is said and what is implicated. The common example of Gricean implicature is complimenting the handwriting of an academic job candidate in a letter of recommendation: what is said (that a candidate has good handwriting) is different from what is implicated (that this is a very weak job candidate). According to Grice:

> A man who, by (in, when) saying (or in making as if to say) that p has implicated that q, may be said to have conversationally implicated that q, provided that (1) he is to be presumed to be observing the conversational maxims, or at least the Cooperative Principle; (2) the supposition that he is aware that, or thinks that, q is required in order to make his saying or making as if to say p (or doing so in those terms) consistent with this presumption; and (3) the speaker thinks (and would expect the hearer to think that the speaker thinks) that it is within the competence of the hearer to work out, or grasp intuitively, that the supposition mentioned in (2) is required. (1975, 30–31)

Put very roughly, S conversationally implies q when saying p when: (1) S implicates q in saying p, (2) S is presumed to be following a principle of conversational cooperation, (3) the supposition that S thinks that q is required to maintain (2), and (4) S thinks the hearer will be able to infer (3).

For my purposes, (3) is the most important point of difference between positive biased evaluative descriptions and instances of traditional Gricean implicature. Gricean implicature requires specific intention on the part of the utterer. Since utterers of positive biased evaluative descriptions do not intend to communicate negative content, positive biased evaluative descriptions do not strictly conform to the letter of Gricean implicature. When Biden called Obama *articulate*, he did not mean to imply anything less than charitable (unlike, for example, 'He has good handwriting'.) Nor are positive biased evaluative descriptions immediately understood or conceptualized as negative even by their audiences.

That makes positive biased evaluative descriptions distinctively pernicious: we are unlikely to realize that even our positive perceptions of people are shaped by implicit bias. The implicature might lurk in the background if one searches for it ('This letter for Joe says that he is the next David Lewis, but this letter for similarly qualified Jane says she is very hardworking'), but many audiences for biased evaluative descriptions will not be consciously aware of the more positive contrast class for diminutive biased evaluative descriptions.

Biased evaluative descriptions also do not easily fit into Jennifer Saul's (2002) expansion of Grice's project to include utterer-implicature and audience-implicature. The general idea of her framework is that there can be speaker meaning that is neither said nor implicated. Biased evaluative descriptions do not fit into this expanded framework since audiences do not necessarily pick up that a description is biased in the relevant way, and utterers do not necessarily intend to communicate negative or biased content. To better understand why, suppose that Joe and Jeffrey are archaic but well-intentioned country club buddies, and Jeffrey agrees with Joe that their African American caddie is *clean-cut*. Neither person is explicitly aware that the description is a diminutive biased evaluative description. Even though Jeffrey, the audience, does not recognize that the description is diminutively racist in context, *clean cut* in this context is obviously a biased evaluative description. Saul also explores the notion of 'unmeant conversational implicatures' (2002: 237–38). Unmeant conversational implicatures are those in which the utterer conversationally implies something that she does not mean. For similar reasons to those already discussed, biased evaluative descriptions do not count as unmeant conversational implicatures either.

Speakers may, of course, communicate negative content without intending to do so. (Because I take the negativity of biased evaluative descriptions to be at the level of pragmatics, I do not explore the option of accounting for biased evaluative descriptions in terms of natural meaning.) Users of diminutive biased evaluative descriptions sometimes fall into this category of unintentional meaning communicators: the well-meaning letter-writer might not think carefully about why she describes Jane as *hardworking* but Joe as *talented*, even while intending to write them both equally strong references. A well-meaning sports announcer might describe Jayvon as *burly* while describing his comparatively light-skinned fellow basketball player as *cunning*. (Indeed, Steven Foy and Rashawn Ray [2019] find stunning differences in terms applied to darker-skinner players by sports announcers.) And Biden meant well in describing Obama as *articulate* and *clean*, even though the dimness of the praise was evident to many other ears. Biden was not attempting to snipe at Obama: to him, *articulate* was to be heard as a genuine compliment by the audience. And the audience does not necessarily hear a negatively valenced implicature, regardless of its intention. ('But he *is* articulate!' is a common refrain utilized in defending the purported positivity of such a description.) Some biased evaluative descriptions might be unintentional instances of what Ishani Maitra (2012) calls 'subordinating speech': speech that subordinates members of marginalized social identities.

Whether biased evaluative descriptions can be made to conform to the spirit of Gricean implicature, or whether traditional Gricean implicature can be expanded to include biased evaluative descriptions, are complex matters. I do not wish to delve further into Gricean speech act theory in this limited space. The goal of this discussion is to show that traditional Gricean implicature does not entirely account for biased evaluative descriptions, even if the framework might be expanded to accommodate them in various ways. Biased evaluative descriptions do not naturally fit into existing taxonomies of similar linguistic phenomena, though some existing concepts might be stretched to include them.

3. Challenges to the Identifiability of Biased Evaluative Descriptions

There are a number of non-linguistic challenges to evaluating whether or not descriptions count as the phenomenon I am describing. First, implicit bias differs between utterers, contexts, and cultures: one person's biased evaluative description is another person's unbiased evaluation. For some, *fun* may be the very highest form of praise given to an academic talk, regardless of the speaker. Some people might have labeled Buttigieg *traditional* no matter what his sexual orientation. It is also likely that such perceptions and labels are highly manipulable. Raising the trait to salience for every political candidate ('On a scale of 1–10, how traditional would you call Pete Buttigieg, Joseph Biden, and Elizabeth Warren, respectively?') would likely increase a description's stability across dominant and subordinate social categories.

Cultural variability also poses a challenge to the identifiability of biased evaluative descriptions. Strength of praise is highly culturally variable and culturally dependent. 'Hard-working' might be the highest form of praise in one culture but not in another. Many languages and dialects have explicitly gendered or racialized compliments. The meanings of the same descriptions differ widely between high-context and low-context cultures. Even within academia, differences in effusiveness between British and American letters of reference are extreme to the point of being widely parodied (Birch 2016).

Intersectionality poses another distinctive challenge to the discernability of particular stereotypes behind biased speech. *Intersectionality*, a concept that originates with Kimberlé Crenshaw, captures the idea that multiple axes of social oppression intersect and interact: 'Consider an analogy to traffic in an intersection, coming and going in all four directions. Discrimination, like traffic through an intersection, may flow in one direction, and it may flow in another. If an accident happens in an intersection, it can be caused by cars traveling from any number of directions and, sometimes, from all of them. Similarly if a black woman is harmed because she is in the intersection, her injury could result from sex discrimination or race discrimination [or both]' (Crenshaw 1989: 149).

The general idea is that members of multiple oppressed social categories—Black women, for example—suffer from 'intersections' of social oppression that are distinct from those of people who are Black and people who are women. Dimensions of oppression mix and interact in ways that add up to greater than the sum of their parts.

A person's membership in an intersectional social category poses a challenge to the identifiability of biased evaluative descriptions because it is unclear whether a description is applied to a person as a member of a unitary oppressed social category, as a member of an intersectional category, or both. Referring to a Black woman as *calm and professional* in a letter of reference, for example, might be a counter-stereotypical biased evaluative description aimed at counteracting the stereotype of Black women as angry and difficult. But it might also be aimed at the stereotype of Black people as angry, or at the stereotype of women as emotional. The use of the term may be conceptually overdetermined, stemming from all three of these stereotypes.

How to understand biased evaluative descriptions with intersectional targets partially depends on how we understand intersectionality. Here I adopt the view set forth in Bernstein (2020), in which I argue that intersectional social categories are best understood as explanatorily unified and explanatorily prior to their constituents. The idea is that intersectional categories like *Black womanhood* back explanations better than unitary social categories like *Blackness* and *womanhood*. Intersectional categories are explanatorily unified because explanations of Blackness cannot be divorced from explanations of womanhood within an intersectional category. Explanations stemming from intersectional categories are more informative and more powerful than explanations exclusively involving the individual identity constituents. (For another metaphysical account of intersectionality, see Jorba and Rodó-de-Zárate [2019]).

Intersectionality creates complexities for counterfactual evaluation of biased evaluative descriptions, since in these cases we must evaluate several alternative worlds in determining the causal roots of the term. Suppose that we are trying to figure out whether a Black woman is labeled *calm and professional* in a letter of reference due to her membership in oppressed social categories. Then we must consider several alternatives. In the world in which the person is neither Black nor a woman, the person in question would not likely be labeled *calm and professional*. But this leaves open whether she might have been labeled as such just for being a woman, or just for being Black. Whether or not she is so labeled because she belongs to the unified intersectional category *Black woman* is underdetermined by the evidence.

4. What is to be Done?

The fact that implicit bias is opaque to introspection poses an extra practical challenge to identifying biased evaluative descriptions. Many of us cannot and will not recognize implicit bias in ourselves, let alone in others. Whether or not we can be held morally responsible for implicit bias itself is a deep puzzle that has only recently gotten significant discussion in the literature. (See Zheng [2016] for a discussion of such issues).

That there are psychological and epistemic barriers to identifying positive biased evaluative descriptions does not mean that we should not try. Recognizing their occurrences and patterns of use is morally important for recognizing discrimination in a variety of social contexts. Spreading the word about biased evaluative descriptions is a promising strategy for ameliorating them.[4] I have

[4] An anonymous referee astutely points out that raising awareness of humblebragging and virtue signaling have helped in ameliorating both phenomena. Perhaps raising awareness about biased evaluative descriptions will help in a similar way.

already mentioned the ubiquitous use of biased evaluative descriptions in letters of recommendation: one is several times more likely to find descriptions like *caring*, *compassionate*, *hardworking*, *conscientious*, *dependable*, *warm*, and *helpful* in letters of reference for women than for men. Biased evaluative descriptions infuse formal and informal discussions about job candidates: it is common to hear well-intentioned but cringe-worthy terms applied to women candidates in addition to or even in lieu of discussions of their research-- *energetic* is one that crops up more often than not. Biased evaluative descriptions are a pervasive feature of political discourse, and are often applied to women candidates and candidates of color. These descriptions reflect discriminatory evaluative judgments that rob their targets of the unqualified, straightforwardly positive evaluations they would receive were they not members of a minority social group. The public application of these terms also robs their targets of the rewards that are causally downstream of unqualifiedly positive evaluations—for example, academic jobs and political offices.

Since biased evaluative descriptions are often vehicles of implicit bias and benevolent discrimination, a natural assumption is that one should never use them, or strive to use the same sorts of descriptions for everyone across the board. But the issue is more complex than these initial treatments suggest.

Counternormative and counterstereotypical biased evaluative descriptions can be effective tools for combating sexism and discrimination. Consider a primary school teacher who, aware of the stereotype of girls as less mathematically talented than boys, makes an extra effort to publicly label them as talented and naturally able. This case satisfies the loose definition of a biased evaluative description: it would not be applied to the students unless they were girls, and the description broadly interacts with a negative stereotype about girls' mathematical talent, even when used intentionally. In this sort of case, careful use of these terms can help to combat negative stereotypes. And it is certainly possible for one to use biased evaluative descriptions thoughtfully—for example, when a job candidate's level of energy for the job might be a particular selling point. Because biased evaluative descriptions do sometimes line up with good traits that should be raised to salience, a blanket recommendation against them is too simplistic.

It is tempting to hold that the solution to the unpleasant downstream effects of biased evaluative descriptions is just to level the evaluative field—to ensure that one is applying the same sorts of descriptions to everyone across the board. But this sort of strategy can go wrong in several ways. First, some biased evaluative descriptions describe traits that are contextually irrelevant. For example, academic letters of recommendation for women tend to positively discuss their physical appearance at a much higher rate than for men. The solution is not to add a discussion of physical appearance in letters of recommendation for men. Rather, the solution is not to discuss features of candidates that are irrelevant to their ability to do their jobs—and to omit these sorts of discussions across the board.

A second problem with leveling the evaluative field is that the same expression can convey and elicit different meanings when applied to different people. *Forceful* might be a positive trait descriptor for a white male philosopher but might elicit negative stereotypes when applied to a Black woman philosopher. Similarly, *traditional* evokes a very different stereotype when applied to Pete Buttigieg as when it is applied to Mike Pence.

A third problem with the strategy of leveling the field is that it reinforces dominant professional values that might be better called into question. In some cases, redefining values might be a better goal than buttressing them. For example, suppose that philosophers particularly value dispassionate ahistoricity as a feature of intellectual work and tend to compliment certain sorts of philosophers on this virtue over others. Rather than compliment more sorts of philosophers on doing this kind of work, the solution might be to call into question why such a value continues to be upheld. Kristie Dotson (2013) complains of a 'culture of legitimation' in academic philosophy that reinforces its own disciplinary and methodological boundaries: 'By relying upon, a presumably, commonly held set of normative, historical precedents, the question of how a given paper is philosophy betrays a value placed on performances and/or narratives of legitimation. Legitimation, here, refers to practices and processes aimed at judging whether some belief, practice, and/or process conforms to accepted standards and patterns, i.e. justifying norms. A culture of justification, then, on my account, takes legitimation to be the penultimate vetting process, where legitimation is but one kind of vetting process among many' (2013: 5).

Vetting processes in academic philosophy, including letters of reference, continue to uphold and reinforce disciplinary norms and values that should be revised. In short, if the playing field improperly favors the dominant group, leveling the playing field is unfair and continues to legitimize the processes by which unfairness is generated. Attending to the underlying dogmas of long-held disciplinary norms, social structures, and philosophical methodologies is a better way forward.

5. Conclusion

Biased evaluative descriptions are terms inflected with implicit bias whose application is counterfactually unstable across dominant and subordinate social groups. Purportedly positive biased evaluative descriptions play roles in social oppression in a variety of contexts, from letters of recommendation to national politics. Learning to recognize biased evaluative descriptions and monitor use of them is an important means for combating implicit bias and social injustice. Through this essay, I hope to have opened an avenue into philosophical investigation of well-intentioned discriminatory speech—how it works, when it occurs, and what its consequences are. The road to hell is paved with good intentions.

SARA BERNSTEIN
UNIVERSITY OF NOTRE DAME
sbernste@nd.edu

References

Adichie, Chimamanda. (2013) *Americanah*. New York: Knopf.

Allen, Samantha. (2017) 'Laverne Cox and the Politics of Transgender Beauty'. *Daily Beast, February* 16, 2017. https://www.thedailybeast.com/laverne-cox-and-the-politics-of-transgender-beauty.

Bernstein, Sara. (2020) 'The Metaphysics of Intersectionality'. *Philosophical Studies,* 177, 321–35.

Bernstein, Sara. (manuscript) 'Countersocial Counterfactuals'.

Bingeman, Emily. (manuscript) 'The Risks of Praise'.

Birch, Jonathan. (2016) 'Academic letters of recommendation: a guide for the perplexed'. Facebook, December 7, 2016. https://www.facebook.com/jonathan.birch.75054/posts/10102039433690700?pnref=story.

Bolinger, Renée Jorgensen. (2017) 'The Pragmatics of Slurs'. *Noûs,* 51, 439–62.

Crenshaw, Kimberlé. (1989) 'Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics'. *University of Chicago Legal Forum*, 1989, article 8. https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1052&context=uclf.

Daly, Helen. (2018) 'On Insults'. *Journal of the American Philosophical Association*, 4, 510–24.

Delston, Jill B. (2021) 'The Ethics and Politics of Microaffirmations'. *Philosophy of Management,* 20, 411–29.

Dotson, Kristie. (2013) 'How Is This Paper Philosophy?' *Comparative Philosophy,* 3, 3–29.

Downs, Jim. (2019) 'Queer Like Pete'. *Slate, November* 25, 2019. https://slate.com/human-interest/2019/11/pete-buttigieg-gay-archetype-best-little-boy.html.

Foy, Steven, and Rashawn Ray. (2019) 'Skin in the Game: Colorism and the Subtle Operation of Stereotypes in Men's College Basketball'. *American Journal of Sociology,* 125, 730–85.

Frye, Marilyn. (1983) 'Oppression'. In *The Politics of Reality: Essays in Feminist Theory* (Trumansburg: Crossing Press), 1–16.

Glick, Peter *et al.* (2000) 'Beyond Prejudice as Simple Antipathy: Hostile and Benevolent Sexism across Cultures'. *Journal of Personality and Social Psychology,* 79, 763–75.

Glick, Peter, and Susan T. Fiske. (1996) 'The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism'. *Journal of Personality and Social Psychology,* 70, 491–512.

Grice, Paul. (1989) 'Logic and Conversation'. In *Studies in the Way of Words* (Cambridge, MA: Harvard University Press), 22–40. Grice's lecture was originally delivered 1967 and published in 1975.

Hideg, Ivona, and D. Lance Ferris. (2016). 'The Compassionate Sexist? How Benevolent Sexism Promotes and Undermines Gender Equality in the Workplace'. *Journal of Personality and Social Psychology*, 111, 706–27.

Hirschberg, Julia Bell. (1985) "A Theory of Scalar Implicature (Natural Languages, Pragmatics, Inference" (PhD diss. University of Pennsylvania, 1985).

Hom, Christopher. (2010) 'Pejoratives'. *Philosophy Compass* 5, 164–85.

Jones, Abigail. (2010) 'Do Jewish Lawyers Really Do It Better?' Forward, September 28, 2010. https://forward.com/schmooze/131666/do-jewish-lawyers-really-do-it-better/.

Jorba, Marta, and Maria Rodó-de-Zárate. (2019) 'Beyond Mutual Constitution: The Properties Framework for Intersectionality Studies'. *Signs: Journal of Women in Culture and Society*, 45, 175–200.

Jost, John T, Laurie A. Rudman, Irene V. Blair, Dana R. Carney, Nilanjana Dasgupta, Jack Glaser, and Curtis D. Hardin. (2009) 'The Existence of Implicit Bias Is Beyond Reasonable Doubt: A Refutation of Ideological and Methodological Objections and Executive Summary of Ten Studies That No Manager Should Ignore' *Research in Organizational Behavior*, 29, 39–69.

Maitra, Ishani. (2012) 'Subordinating Speech'. In Mary Kate McGowan and Ishani Maitra (eds.), *Speech and Harm: Controversies Over Free Speech* (Oxford: Oxford University Press), 94–120.

Marques, Teresa, and Manuel García-Carpintero. (2014) 'Disagreement about Taste: Commonality Presuppositions and Coordination'. *Australasian Journal of Philosophy,* 92, 701–23.

McGrath, James. (2019) 'Not All Autistic People Are Good at Maths and Science—Despite the Stereotypes'. *The Conversation, April* 3, 2019. https://theconversation.com/not-all-autistic-people-are-good-at-maths-and-science-despite-the-stereotypes-114128.

Meyer, Meredith, Andrei Cimpian, and Sarah-Jane Leslie (2015). 'Women Are Underrepresented in Fields Where Success Is Believed to Require Brilliance'. *Frontiers in Psychology*, 6, article 235. https://doi.org/10.3389/fpsyg.2015.00235.

NPR. (2007) 'Political Experts Weigh in on Biden Gaffe'. NPR News, February 2, 2007. https://www.npr.org/templates/story/story.php?storyId=7127782.

Popa-Wyatt, Mihaela, and Jeremy L. Wyatt. (2018) 'Slurs, Roles and Power'. *Philosophical Studies,* 175, 2879–2906.

Rett, Jessica. (2020) 'Manner Implicatures and How to Spot Them'. *International Review of Pragmatics,* 12, 44–79.

Saul, Jennifer. (2002) 'Speaker Meaning, What Is Said, and What Is Implicated'. *Noûs,* 36, 228–48.

Schiavenza, Matt. (2018) 'Silicon Valley's Forgotten Minority'. *New Republic, January* 11, 2018. https://newrepublic.com/article/146587/silicon-valleys-forgotten-minority.

Schlenker, Philippe. (2012) 'Maximize Presupposition and Gricean Reasoning' *Natural Language Semantics*, 20, 391–429.

Sennet, Adam, and David Copp. (2015) 'What Kind of a Mistake Is It to Use a Slur?' *Philosophical Studies*, 172, 1079–1104.

Smith, David. (2019) 'How the Other Halves Live: Candidates' Spouses Show Modern American Family'. *The Guardian*, December 28, 2019. https://www.theguardian.com/us-news/2019/dec/27/democratic-candidates-spouses-partners-modern-american-family.

Trix, Frances, and Carolyn Psenka. (2003) 'Exploring the Color of Glass: Letters of Recommendation for Female and Male Medical Faculty'. *Discourse & Society,* 14, 191–220.

Zheng, Robin. (2016) 'Attributability, Accountability, and Implicit Bias'. In Michael Brownstein and Jennifer Saul (eds.), *Implicit Bias and Philosophy,* vol. 2, *Moral Responsibility, Structural Injustice, and Ethics* (New York: Oxford University Press), 62–89.