Biased Evaluative Descriptions[1]
Sara Bernstein

Abstract: This paper identifies a type of linguistic phenomenon new to feminist philosophy of language: biased evaluative descriptions. *Biased evaluative descriptions* (BEDs) are descriptions whose well-intended positive surface meanings are inflected with implicitly biased content. Biased evaluative descriptions are characterized by three main features: (i) they have roots in implicit bias or benevolent sexism, (ii) their application is counterfactually unstable across dominant and subordinate social groups, and (iii) 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 framework of traditional Gricean implicature cannot account for BEDs. I discuss some challenges to the distinctiveness and evaluability of BEDs, including intersectional social identities. I conclude by discussing the social significance and moral status of BEDs. Identifying BEDs 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 and bright and clean and a nice-looking guy."[2] 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.[3] Trans actress Laverne Cox is often described as "feminine" by well-meaning celebrity journalists. Though "articulate", "clean", "traditional", and "feminine" 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

[2] https://www.npr.org/templates/transcript/transcript.php?storyId=7127782

[3] For example, https://www.theguardian.com/us-news/2019/dec/27/democratic-candidates-spouses-partners-modern-american-family.

apparently positive surface meanings are inflected with implicit bias or benevolent discrimination, will be the focus of this paper.

The primary goals of the discussion will be to characterize the phenomenon, and to make headway in diagnosing the linguistic force and moral significance of such descriptions. The secondary goal will be 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.

I proceed in several stages. In section 1, I describe biased evaluative descriptions (hereafter: BEDs) and give several sorts of examples of them. I suggest that BEDs are characterized by three main features: they often have roots in implicit bias or benevolent sexism, their application is counterfactually unstable across dominant and subordinate social groups, and they encode stereotypes. I also describe what is morally problematic about BEDs. In section 2, I explore where BEDs fall in existing taxonomies of related linguistic phenomena. I distinguish BEDs from similar linguistic concepts, including straightforward compliments, backhanded compliments, euphemisms, slurs, insults, epithets, and dog-whistles. I discuss Gricean implicature and speech act theory in relation to BEDs. In section 3, I address obstacles and objections to the distinctiveness and identifiability of BEDs, including challenges having to do with intersectional social identities inhabited by their targets. In section 4, I discuss the moral status of BEDs. The use of BEDs is not always morally problematic, I suggest, but awareness of their usage and occurrence is important for differentiating their good and bad uses. Recognizing the occurrence of BEDs is illuminating in a variety of social contexts, from the very general and broad (political speeches) to the very particular and small (letters of recommendation and bias in academic hiring). I conclude with some remarks on what can be done to remedy the situation.

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, as when women athletes are asked by journalists 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 will be relatively narrow, some of my remarks will 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.

Call the examples with which I began *positive* biased evaluative descriptions. Positive BEDs are intended to be complimentary. They are to be contrasted with negative BEDs, 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 since I am interested in well-intentioned sexist speech rather than speech backed by overt sexism and discrimination, I will focus primarily on positive BEDs.[4]

BEDs encompass a broad but overlapping range of phenomena. Some BEDs are primarily *counterstereotypical*: their usage is intended to negate stereotypes associated with particular social groups. The examples with which I began are counterstereotypical BEDs. 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 "feminine", the press intends to counter the stereotype of trans women as not entirely women.[5] A similar example comes from a hijab-wearing Muslim student of mine who complains 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

---

[4] Many of my points additionally apply to intentionally negative BEDs. Thanks to a referee for pointing this out.

[5] An overlapping phenomenon is the biased compliment, e.g., "Yasss king!" uttered by cis persons towards trans men as a well-intentioned affirmation of the latter's masculinity.

about me if I did not wear the hijab."[6]

Other BEDs are *counternormative*. Users of counternormative BEDs 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*:

> 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, pp. 180-181)

In this context, "beautiful" functions as a counternormative BED because it 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. Like other positive BEDs, the compliment is well-intentioned. But such a compliment has its roots in a form of benevolent racism according to which all women belonging to a certain racial category count as beautiful—an essentializing, oversimplified claim. (I discuss the concept of benevolent racism below in section 1b.)

Other BEDs are primarily *diminutive*: even while they are intended to be straightforwardly complimentary, their positive force is diminished due to comparison class of descriptions that are applied to members of dominant social categories. For example, the

---

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

adjective most often used to compliment my professional philosophy talks is "fun".[7] My talks *do* tend to be fun: they contain interesting examples and snappy jokes, and they are constructed with clarity and accessibility in mind. But "fun" diminishes the other, more professionally valuable aspects of the talk (clarity, explanatory power, creativity, intellectual depth) in favor of one considered less important to professional prestige. Even apparently straightforward descriptions such as "accomplished" and "professional" can count as diminutive BEDs, depending on context and utterer.

Diminutive BEDs are pervasive in letters of recommendation for women. A prominent study on differences between letters of recommendation for male and female medical faculty candidates found a significant percentage increase in "grindstone adjectives", or adjectives used to describe being hard-working, in letters for women.[8] Aside from "hard-working", typical examples of grindstone adjectives include "conscientious" and "diligent". Being hard-working is obviously a good trait for any faculty member, but it 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. 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 white men.[9]

Many BEDs 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.[10] 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 BEDs can span categories doesn't diminish the explanatory power of

---

[7] I hasten to add that I am not at all offended by the "fun" label, and in fact pride myself on giving lively, engaging talks. I merely aim to draw attention to the frequency of the label as applied to my talks in contrast with the apparent sparsity with which it is applied to similar talks given by male colleagues.

[8] See Trix, F., & Psenka, C. (2003).

[9] See Meyer, Meredith; Cimpian, Andrei & Leslie, Sarah-Jane (2015).

[10] It has been suggested to me that perhaps all BEDs might turn out to be diminutive. Due to space constraints, I do not argue that this is the case in this paper. Even if there is a unified phenomenon among the three types of BEDs, disentangling them is theoretically and explanatorily useful.

the categories themselves, since the categories help us understand BEDs and more carefully identify their effects.

Finally, while my focus is BEDs that are applied to marginalized groups, there are numerous descriptive utterances that count as BEDs directed at people in dominant groups. Consider one friend saying to another of a party invitee: "He's white, but don't worry, he's cool."[11] This BED exhibits counterfactual instability 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.

Now that we have a general grasp of the examples and the different sorts of categories under which they fall, we can draw a few more contours around them. Most BEDs have three major elements: (i) they involve implicit bias or benevolent discriminatory attitudes, (ii) they are counterfactually unstable across subordinate and dominant social identities, and (iii) they involve or encode stereotypes. I'll briefly discuss each element in turn.

a. Implicit Bias and Benevolent Discriminatory Attitudes

Many positive BEDs are a product 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.[12]

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 BEDs signals such biases. This point is important for zeroing in on the sort of apparently positive descriptions that are the subject of my investigation.

Glick and Fiske (1996) draw a distinction between *hostile sexism* and *benevolent sexism*. Hostile sexism targeted at women is undergirded by explicitly negative and reductionist attitudes towards women—think Rush Limbaugh, Ann Coulter, and *The Handmaid's Tale*. Benevolent

---

[11] Thanks to a member of the audience in the Rutgers Feminist Philosophy Group for this example.
[12] See, e.g., Jost et. al (2009).

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.[13] 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) (Glick & Fiske, 1996, p. 491). [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, p. 763)."

In other words, many instantiations of benevolent sexism are intended to be positive or supportive of women, but in fact reinforce their subordinate social status. Benevolent sexist attitudes include paternalistic attitudes towards women's emotional fragility (e.g. women must be emotionally protected because they are "more pure"), essentialist views about women's supposed goodness (e.g. women are naturally more compassionate and nurturing), and reductive views about women's dispositions and abilities (e.g. women are less aggressive than men, which makes them less suited to aggressive q&a in academic philosophy.)

Benevolent sexism is a progenitor of many well-intentioned positive BEDs. Benevolent sexism is pervasive among well-meaning self-identified allies of women and minorities. It is even the case that holders of benevolent sexist attitudes are more likely to support equal opportunity policies in the workplace.[14] Users of positive BEDs intend to be supportive and well-meaning, and often work towards genuine social good. Benevolent sexism is also pervasive. A significant amount of 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

---

[13] 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.

[14] Hideg, I., & Ferris, D. L. (2016). The compassionate sexist? How benevolent sexism promotes and undermines gender equality in the workplace. Journal of Personality and Social Psychology, 111(5), 706–727. https://doi.org/10.1037/pspi0000072

"naturally" more compassionate and reasonable.

There are parallel 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.[15] People on the autism spectrum complain of being stereotyped as mathematical or scientific savants.[16] Jewish lawyers complain that they are hired because they are thought to automatically be more effective at practicing law than others.[17]

The manner and extent to which a BED is caused by implicit bias varies by type of BED. Counterstereotypical and counternormative BEDs often involve a conscious effort to counteract stereotypes and norms, as with Kimberly in the fictional Adichie case. Diminutive BEDs often involve unencumbered bias that is buried much deeper. In the case of a woman's talk labelled "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.[18] An entire paper could be written about these variations, and I cannot do justice to the topic here. I hope that it is the beginning of the investigation into these phenomena.

b. Counterfactual Instability Across Subordinate and Dominant Social Groups

Many positive BEDs express an evaluation of a person that would not be applied were a particular sort of subordinate social identity owing to gender, race, ability, socioeconomic status, etc. not occupied by the person being evaluated. I call this property of BEDs *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. Entertaining such *countersocial counterfactuals*, counterfactuals that run contrary to social fact, is an intuitive

---

[15] https://newrepublic.com/article/146587/silicon-valleys-forgotten-minority

[16] https://theconversation.com/not-all-autistic-people-are-good-at-maths-and-science-despite-the-stereotypes-114128

[17] https://forward.com/schmooze/131666/do-jewish-lawyers-really-do-it-better/

[18] Thanks to a referee for articulating these differences in this way.

method of identifying BEDs.[19]

While we often seem to entertain countersocial counterfactuals fairly easily ("Would he have said that if I was a man?"), what we are implicitly doing 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 counterfactual evaluation ignores the details of social constructions of race, and its associated ideologies. The conceptually rich 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. Similarly for other such countersocial counterfactuals, which are not just a matter of swapping out physical traits or parts. It is hard to give a good story about how we manage to evaluate countersocial counterfactuals so reliably, and I will not try to give a story here. I will assume in the following discussion, however, that we can generally agree on how to evaluate them.

Being 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 "hard-working" in letters of recommendation presumably *are* hard-working.

"What's wrong with saying she's hard-working?" is a common refrain among those who use this and other grindstone adjectives. What's wrong is that "hard-working", "energetic", and many similar positive BEDs 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 BEDs: if a person 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

---

[19] See my "Countersocial Counterfactuals" (manuscript).

descriptions indicating traits of less professional value (for example, those alluding to hard work like "persistent"), then the descriptions are likely to exhibit bias in those instances, *ceteris paribus*.[20]

c. Encoding of Stereotypes

Finally, most BEDs encode stereotypes. Some do so fairly explicitly. Hearing Buttigieg described as "traditional" immediately calls up mental images of the contrast class of "non-traditional" 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 BEDs encode stereotypes more covertly. In the context of a letter of recommendation, a "hard-working" 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. While these descriptions are intended as compliments, they instead reinforce stereotypes by encoding them. They at once grant faint praise on the target, while essentializing and stereotyping extant members of the target's social category. This is one way that users of well-intentioned BEDs go awry: in encoding the stereotype, they reinforce the pernicious norms and expectations that they intend to dissolve with the positive evaluation.

d. What is Problematic about BEDs?

The above discussion helps get a handle on the phenomenon of BEDs, but it does not entirely capture what is harmful about them. I have alluded to the idea that BEDs are problematic because they figure into particular patterns of use. 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 student papers are described differently than papers by colleagues.[21]

BEDs are harmful in several ways. First, BEDs can create or reinforce low expectations

---

[20] Obviously professional values vary by context. "Hard-working" and "persistent" in the mining industry will have a different implicatures than "hard-working" in academic philosophy.

[21] Many thanks to an insightful referee for this point.

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. BEDs 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, thoughts, and expectations of social groups, BEDs play a role in negatively shaping concepts of already-marginalized social groups.

Second, BEDs can lead to harms subsequent to their use.[22] One sort of harm is that 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.

BEDs can also do harm through their encoding of stereotypes. As I suggested above, some BEDs pit their subjects against fellow members of marginalized social groups. For example, calling a hijab-wearing Muslim woman "open-minded" calls to mind contrastive stereotypes of dogmatic hijab-wearing women. On the other hand, indiscriminate application of terms, such as "beautiful" for all Black women, deprives the 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), 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.[23] Stereotypes socially punish their subjects whether or not the

---

[22] Emily Bingeman has work in preparation on the moral risks of these forms of praise.

[23] https://slate.com/human-interest/2019/11/pete-buttigieg-gay-archetype-best-little-boy.html

subjects conform to expectations.

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

2. BEDs and Existing Linguistic Frameworks

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

BEDs are not straightforward compliments. Straightforward compliments are uncomplicatedly positive, e.g., "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 BEDs 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 BEDs via speaker intention: positive BEDs are intended to be positive, whereas backhanded compliments are not. BEDs 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; BEDs are not.

BEDs 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.[24] For similar reasons, BEDs are not pejoratives, which convey intentionally negative content.[25] Utterers of positive BEDs do not explicitly endorse negative or pejorative content. BEDs 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 Jews "cosmopolitan". Users of positive BEDs generally do not intend to encode negative or stereotypical content, and the audience for such descriptions is public.

Positive BEDs are not straightforward insults, which intentionally communicate negative information or lack of respect about the target. It might turn out that some BEDs are unintentional or non-straightforward insults, if there are such things.[26] Well-intentioned users of BEDs intend to communicate positive features of the target.

BEDs are not euphemisms, which indirectly name a trait or a cluster of traits. For example, "electability" in 2020 American democratic primary election encodes male traits. Racist media directed towards Meghan Markle that labels her "exotic" encodes African-American traits. But BEDs straightforwardly attribute traits rather than shrouding them in euphemisms.

Some BEDs have features in common with scalar implicature.[27] 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.[28] 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 BEDs 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 BEDs, choosing one sort of term over another for a marginalized group is closely related to why such BEDs are problematic. But while scalar implicature can help illuminate the mechanisms behind some BEDs, scalar implicature

---

[24] See, e.g., Bolinger (2017) for a pragmatic account of slurs. See Popa-Wyatt and Wyatt (2018) for an account of slurs that incorporates dominant and submissive social roles.

[25] For recent accounts of pejoratives, see Hom (2010), Sennet and Copp (2015), and Marques and García-Carpintero (2020).

[26] This will depend partly on one's view of insults. See, e.g., Daly (2018) for a view of insults as expressions of lack of due regard.

[27] Thanks to a helpful referee for drawing attention to this similarity.

[28] For a good discussion of scalar implicature and related phenomena, see Rett (2020). For a helpful discussion of maximizing presupposition, a phenomenon closely related to scalar implicature, see Schlenkler (2012).

does not perfectly align with the phenomenon, for two reasons. First, not all BEDs fall on a single informational spectrum. For example, "articulate" isn't 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 BEDs.

It might be tempting to try to explain BEDs in terms of Grice's (1975) 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 than 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, pp. 30-31)

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

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

That makes positive BEDs 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 hard-working"), but many audiences for biased evaluative descriptions will not be consciously aware of the more positive contrast class for diminutive BEDs.

BEDs 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. BEDs do not fit into this expanded framework since audiences do not necessarily pick up that a description is a BED, 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 BED. Even though Jeffrey, the audience, does not recognize that the description is diminutively racist in context, "clean cut" in this context is obviously a BED.[29]

Speakers may, of course, communicate negative content without intending to do so.[30] Users of diminutive BEDs sometimes fall into this category: the well-meaning letter-writer might not think carefully about why she describes Jane as hard-working 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.[31] 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.)

Whether BEDs can be made to conform to the spirit of Gricean implicature, or whether

---

[29] Saul (2002) also explores the notion of "unmeant conversational implicatures". Unmeant conversational implicatures are those in which the utterer conversationally implies something that she does not mean. For similar reasons to those already discussed, BEDs do not count as unmeant conversational implicatures, either.

[30] Since I take the negativity of BEDs to be at the level of pragmatics, I do not explore the option of accounting for BEDs in terms of natural meaning.

[31] Indeed, Foy and Ray (2019) find stunning differences in terms applied to darker-skinner players by sports announcers.

traditional Gricean implicature can be expanded to include BEDs, 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 doesn't entirely account for BEDs, even if the framework might be expanded to accommodate them in various ways. BEDs do not naturally fit into existing taxonomies of similar linguistic phenomena.[32]

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 BEDs. First, implicit bias differs between utterers, contexts, and cultures: one person's BED 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 BEDs. 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.[33]

Another distinctive challenge to the discernability of the particular stereotypes that cause BEDs is intersectionality. *Intersectionality*, a concept that originates with Kimberlé Crenshaw (1989), captures the idea that multiple axes of social oppression intersect. According to Crenshaw:

---

[32] Some BEDs might be instances of what Maitra (2012) calls "subordinating speech".

[33] E.g. https://www.facebook.com/jonathan.birch.75054/posts/10102039433690700?pnref=story

"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, p. 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 BEDs 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 BED 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 BED may be conceptually overdetermined, stemming from all three of these stereotypes.

How to understand BEDs with intersectional targets partially depends on how we understand intersectionality. Here I adopt the view set forth in Bernstein (2020), which argues 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.[34]

Intersectionality creates complexities for counterfactual evaluation of BEDs, since in these cases we must evaluate several alternative worlds in determining the roots of a BED.

---

[34] For another metaphysical account of intersectionality, see Jorba & Rodó-de-Zárate (2019).

Suppose that we are trying to figure out whether a Black woman is labelled "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 likely not 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 BEDs. Many of us cannot and will not even recognize the bias in ourselves, let alone in others.[35]

That there are psychological and epistemic barriers to identifying positive BEDs does not mean that we should not try. Recognizing the occurrence and patterns of use of positive BEDs is morally important for recognizing discrimination in a variety of social contexts. Spreading the word about BEDs is a promising strategy for ameliorating them.[36] I have already mentioned the ubiquitous use of BEDs in letters of recommendation: one is several times more likely to find descriptions like "caring", "compassionate," "hard-working," "conscientious," "dependable," "warm," and "helpful" in letters of reference for women than for men.[37] BEDs infuse formal and informal discussions about job candidates: it is common to hear well-intentioned but cringe-worthy BEDs 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.) BEDs are a pervasive feature of political discourse, and are often applied to women candidates and candidates of color. Many BEDs reflect discriminatory evaluative judgments that rob their targets of the unqualified, straightforwardly positive evaluations they would receive were they not members of a minority

_____

[35] 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. How responsible we are for identifying and eliminating positive BEDs will depend on our responsibility for dealing with implicit bias more generally. See Zheng (2016) for a discussion of such issues.

[36] An anonymous referee astutely points out that raising awareness of humblebragging and virtue signaling have helped in ameliorating both phenomena.

[37] See, e.g, Trix, F & Psenka, C (2003) and Madera, JM, Hebl, MR, & Martin, RC (2009).

social group. BEDs also rob their targets of the rewards that are causally downstream of unqualifiedly positive evaluations—for example, academic jobs and political offices.

Since BEDs are often vehicles of implicit bias and benevolent discrimination, a natural assumption is that one should never use them, or strive to level the evaluative field—to use the same sorts of descriptions for everyone across the board. But the issue is more complex than this initial treatment suggests.

Counternormative and counterstereotypical BEDs 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 BED: 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 BEDs can help to combat negative stereotypes. And it is certainly possible for one to use BEDs thoughtfully—for example, when a job candidate's level of energy for the job might be a particular selling point. Since BEDs do sometimes line up with good traits that should be raised to salience, a blanket recommendation against BEDs is too simplistic.

It is tempting to hold that the solution to the unpleasant downstream effects of BEDs 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 BEDs describe traits that are contextually irrelevant. For example, academic letters of recommendation for women tend to positively discuss their grooming at a much higher rate than for men. The solution is not to add a discussion of grooming 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 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 for a white male philosopher to have, 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 "leveling the field" strategy is that doing so 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. The solution might not be to compliment more sorts of philosophers on doing this kind of work, but rather to call into question why such a value continues to be upheld. In her (2012), Kristie Dotson complains of a "culture of legitimation" in academic philosophy that reinforces its own disciplinary and methodological boundaries. She writes:

> "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." (2012, p. 5)

Vetting processes in academic philosophy, presumably 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

This paper has explored and examined the phenomenon of *biased evaluative descriptions*, descriptions 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 BEDs, and monitor use of them, is an important means for combating implicit bias and social injustice. More generally, 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.

References

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

Bernstein, S. (2020) "The Metaphysics of Intersectionality" *Philosophical Studies* 177:2 321-335.

Bernstein, S. (manuscript) "Countersocial Counterfactuals".

Bolinger, R. (2017) "The Pragmatics of Slurs" *Nous,* vol. 51 no. 3: 439-462.

Crenshaw, K. (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*: Vol. 1989, Article 8.

Daly, H. (2018) "On Insults" *Journal of the American Philosophical Association* 4 (4): 510-524.

Dotson, K. (2013) "How is this Paper Philosophy?" *Comparative Philosophy* 3 (1):3-29.

Foy, S. and Ray, R. (2019) "Skin in the Game: Colorism and the Subtle Operation of Stereotypes in Men's College Basketball" *The American Journal of Sociology* Vol.125 (3) 730-785.

Frye, M. (1983). "Oppression" in *The Politics of Reality: Essays in Feminist Theory*. Trumansburg, NY: The Crossing Press.

Glick, P., & Fiske, S. T. (1996). "The ambivalent sexism inventory: Differentiating hostile and benevolent sexism" *Journal of Personality and Social Psychology*, 70, 491–512.

Grice, P. (1975) "Logic and Conversation", in Cole & Morgan 1975: 41–58. Reprinted in Grice 1989: 22–40.

Hirschberg, Julia Bell (1985) "A Theory of Scalar Implicature (Natural Languages, Pragmatics, Inference)" Dissertations available from ProQuest. AAI8603648.

Hom, C. (2010) "Pejoratives" Philosophy Compass 5/2: 164–185, 10.1111/j.1747-9991.2009.00274.x

Maitra, I. (2012). Subordinating Speech. In Mary Kate McGowan Ishani Maitra (ed.), *Speech and Harm: Controversies Over Free Speech*. Oxford: Oxford University Press. pp. 94-120.

Marques, T. & García-Carpintero, M. (2020) "Really expressive presuppositions and how to block them" *Grazer Philosophische Studien* 97 138-158.

Meyer, M, Cimpian, A. & Leslie, S-J. (2015) "Women are underrepresented in fields where success is believed to require brilliance" *Frontiers in Psychology* 6.

Jorba, M. & Rodó-de-Zárate, M. (2019) "Beyond Mutual Constitution: The Properties Framework for Intersectionality Studies." *Signs: Journal of Women in Culture and Society* 45 (1):175-200.

Jost, J.T., Rudman, L.A., Blair, I.V., Carney, D., Dasgupta, N., Glaser, J. & Hardin, C.D. (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.

Popa-Wyatt, M. & Wyatt, J. (2018) "Slurs, roles and power" *Philosophical Studies* 175 (11):2879-2906.

Rett, J. (2020) "Manner implicatures and how to spot them" International Review of Pragmatics, 12(1), 44-79.

Saul, J. (2002) "Speaker meaning, what is said, and what is implicated." *Noûs* 36 (2):228–248.

Schlenker, P. (2012) *Maximize Presupposition* and Gricean reasoning. *Nat Lang Semantics* 20, 391–429.

Sennet, A & Copp, D. (2015) "What kind of a mistake is it to use a slur?" *Philosophical Studies* 172 (4):1079-1104.

Trix, F., & Psenka, C. (2003) "Exploring the color of glass: Letters of recommendation for female and male medical faculty." *Discourse & Society, 14*(2), 191-220.

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