Labeling Unlabeled Identities

Katherine Ritchie
University of California, Irvine

Forthcoming in
New Perspectives on Conceptual Engineering Vol. 3: "Applied Conceptual Engineering"
M.G. Isaac, S. Koch, & K. Scharp (Eds.), Springer

Abstract: There is a longstanding debate about the harms and benefits of appealing to identities and using identity labels in social, political, and academic contexts. While there are potential harms with both labeling lacunae and with using labels, by considering the interplay of power, the interests and epistemic positions of the powerful and marginalized, and the role self-labeling plays, I argue that the introduction and usage of labels for both dominant and marginalized identities can bring important benefits and that strategies can be taken to mitigate harms. This shows, at a minimum, that blanket claims that identity terms are destructive or that identity labels do more bad than good are too strong.

"I'm a bitch, I'm a lover
I'm a child, I'm a mother
I'm a sinner, I'm a saint
I do not feel ashamed"

—Meredith Brooks 'Bitch' from *Blurring the Edges* (1997)

The role identities and identity terms ought to play in policy and education is currently being debated in both academic circles and within the public sphere. Some argue that relying on identities and using identity labels leads to more harm than good. It divides when we need unity; it trains our attention on particularities, when universals ought to be emphasized (e.g., Lilla 2017; Appiah 2018). Advice suggesting that we ought to stop labeling to promote better self-esteem, teaching, and parenting abounds.[1] Further some argue that minimizing labeling can promote a more just world. For example, Leslie argues that reducing the use of labels may help limit prejudice and the extent to which people essentialize groups (2017: 418-420). Others argue that identity talk is needed to understand the ways people are differentially positioned, to reckon with systematic forms of oppression, and to explain differences in lived experience and the distribution of resources (e.g., Young 1990; Alcoff 2006).

[1] A Google search found top results with advice to "stop labeling" or "ditch the labels" in *Psychology Today*, *Forbes,* and many sites with parenting and teaching advice.

There are many ways to intervene in these debates. My focus here is on a question related to language. Think, for a moment, of the expressions that make up the lexicon of a particular language as tools. These tools facilitate discussion, explanation, making salient, and reference. In addition to considering a language as it currently is, we can also imagine alternative ways that a language could be. We can imagine a language with different tools, with more tools, or with fewer. For instance, we can envisage a language that does not have expressions that categorize or label people in ways related to race, gender, or sexual orientation. Or we could imagine one that is more expressively rich and allows for further identities to be picked out. Here I consider the harms and benefits of introducing new identity labels into a language. Considering these will help to shed some light on what identity terms can do and whether or when labels ought to be introduced or eliminated.

As I am using the term, *labels* are expressions that are lexicalized in a language (i.e., are in the lexicon of the language) and which purport to pick out (objects, events, kinds, or…) and categorize. When asking whether a label should be introduced for something, *A*, we are asking whether there are reasons to prefer a language with a compact lexicalized way of picking out *A* to a language without one.[2] We can also consider the effects of various sorts of labels. For instance, what differences are there between labeling with adjectives rather than nouns (Wierzbicka 1986; Markman 1989; Gelman 2003; Ritchie 2021b)? Are more fine- or coarse-grained labels preferable (for a particular purpose or in a particular context)? My aim here is not to answer the many important questions about what sorts of labels to use, but to consider the effects of labels as a broader class by contrasting languages with more limited linguistic resources and those with richer resources in order to argue that linguistic innovation regarding identity language can bring important benefits.

We have labels for many social identities.[3] People are labeled and self-label with *woman*, *Black, American, Republican, queer*, *liberal* and so on. There are also identities for which there are no labels whatsoever (in a given language, *L*). While I cannot populate a list of labels for identities that have none, *genderfluid, nonbinary, cis(gender)*, and *able-bodied* are identities for which labeling lacunae have recently been filled. There are also identities for which there are labels (in *L*) that are infrequently used. For example, in many contexts certain identities are taken to be the default or norm, so much so that

[2] Both nouns and adjectives are labels on this construal given the understanding of lexical categories in Szabó (2015).
[3] In using *identity* here and throughout this paper I have in mind what are sometimes called "collective identities" like racial, gender, social-economic class, and religious identities. I do not mean to be focusing on what is sometimes called "individual identity" which is tied to a particular self-representation.

labeling them might seem odd. For instance, in some contexts it is uncommon to use *white*, *heterosexual*, and *cis*. While my focus is on identities for which there are no labels (in *L*), in analyzing the dynamics of unlabeled identities, appeal to infrequently labeled identities will sometimes be useful in drawing out consequences of introducing and coming to use a label.

To answer the questions about whether and when introducing an identity label can be beneficial, I consider the effects of introducing and applying labels to unlabeled, but existent, identities.[4] I focus on these cases, given the following thinking. If labeling identities generally has negative effects that are not and cannot be outweighed by benefits, that strongly suggests that we should not *proliferate* identity terms. Perhaps given the way our language currently is we should not eliminate all the identity labels that already exist. It might not be possible to do so. It might lead to backlash that would nullify the benefits of a language without identity labels. But, if labeling identities is generally harmful, it is plausible to conclude that we should not add *more* identity labels.

Both sides in debates about identity labels are right in some respects—there are potential harms from using labels and from not using labels. I argue using two case studies that the introduction of labels for previously unlabeled dominant and marginalized identities can bring important epistemic, social, and political benefits and that potential harms can be mitigated. This points to conditions under which introducing a label is positive. It also shows that broad criticisms that labeling is more bad than good or that appealing to identities is always destructive are too strong. Returning to the tool metaphor, it can be useful to have a greater range of tools, but this does not show that any one tool (e.g., some particular label for an identity group) is the best expression to use for all purposes or in every context. Rather, just as a hammer can be useful for some purposes and a wrench can be better for others, identity labels can facilitate explanation, understanding, organizing, and other positive ends, but other sorts of expressions can be useful as well. This claim also allows us to recognize that labels can be used in ways that are harmful (just like hammers!). The call for richer representational resources for identities does not involve a mandate that we use identity talk in all contexts or for all (social, political, epistemic…) purposes.

[4] I focus on consequences here. I take these to be morally significant whether one is a consequentialist or non-consequentialist. In considering identity-talk, consequences are often brought to the fore. For instance, in arguing that identity talk is divisive, divisions have negative ramifications on society, and, therefore, identity talk should be limited or eliminated, one is relying on the consequences of identity talk. One might argue that agents have an interest in self-labeling and having the tools to do so is good in and of itself, without appeal to the consequences of labeling. I won't engage in that project here, but take it to be a welcome addition. I thank an anonymous referee for suggestions on this point.

My focus here is on language and on introducing expressions into a language. These projects fit into the broader project of conceptual engineering understood as at least partially linguistic.[5] There is currently debate about what the target of conceptual engineering is (see Isaac (2021a) and (2021b) for a useful taxonomy and arguments which I draw from here). Some argue its target is representational devices in general without requiring further specification (e.g., Burgess and Plunkett 2013a, 2013b, 2020; Cantalamessa 2019; Nado 2019; Sterken 2020). Others take its target to be something specifically mental (e.g., concepts; see Isaac 2021a, 2021b). Still others take it to target something else, for instance something specifically linguistic (Cappelen 2018). Since my focus is on language, whether this directly fits into a project of conceptual engineering depends on how one construes such projects. On a view on which engineering linguistic labels (i.e., introducing new words) is at least one way to engage in conceptual engineering, then my aims here do fall within conceptual engineering. On a view on which conceptual engineering is about concepts in particular, the project I am engaged in here is a form of what Isaac (2021a) suggests could be called *lexical engineering*.

It is worth noting that even if one construes conceptual engineering as targeting only concepts, one might still hold that lexical engineering is a means to changing concepts. For instance, one might hold that learning new words also involves acquiring or facilitating the use of a concept. I will not take a stand on questions about concept learning or acquisition or other questions about the how language affects cognition here. Instead, I will largely constrain my discussion to language, leaving what implications for concepts are to be drawn (or not) open depending on one's more general views on conceptual engineering and the mind-language interface.

The paper is structured as follows. I begin by arguing that there are unlabeled identities (*I*). While labels can make shared identities more salient, I argue that there is a meaningful notion of identity on which identities can exist without being labeled. Then I consider the effects of failures to label identities (*II*). I consider general effects that result from labeling lacunae as well as more specific effects, in particular harms, that come from failing to label dominant and marginalized identities. Next, I consider the effects that come from labeling, including worries that labels are divisive and that they promote bias and essentialism (*III*). I then turn to my main argument that labeling unlabeled dominant and marginalized identities brings important benefits (*IV*). Finally, I consider lingering worries and

[5] For discussions and overviews of conceptual engineering see Cappelen (2018), Burgess and Plunkett (2013a) and (2013b), Burgess et al. (2020), and Isaac et al. (2022).

conclude (*V*). Labels and labeling can be positive; blanket claims that identity labels are harmful and should be avoided are too strong.

I. The Existence of Unlabeled Identities

In order for the project of examining the effects of introducing labels for unlabeled identities to get off the ground, it must be possible for there to be unlabeled social identities. Whether there can be social identities for which there is no label whatsoever (in a linguistic community in which there are individuals with the identity) is controversial. For instance, Appiah (1996; 2005) argues that identities require labels.[6]

My aim here is not to argue for a complete metaphysics of identities, social groups, or other elements of the social world. Rather, my aim is to make a case that there can be unlabeled identities, by drawing on prominent views in social metaphysics. To begin to make this case, consider the fact that other social entities can exist without being labeled. For instance, Thomasson (2003a, 2003b) and Khalidi (2013) have influentially argued that recessions and racism could exist without our having any concepts or expressions that pick them out. Haslanger (2003: 314) argues that there can be social categories that are not named or explicitly classified in a language. She offers as a possible example children of widows (in a particular culture), which she suggests might involve a shared social position without a label for the category. This provides some prima facie evidence for unlabeled identities, insofar as other social entities are not inherently dependent on linguistic resources. To further bolster the case, let's consider a prominent way of understanding what identities are. (There may be other metaphysical views that also support the existence of unlabeled identities, but I will just focus on this one to offer a concrete illustration).

There is a robust tradition in feminist and critical race theory of understanding identities via an analogy with maps. As Cross argues, "one's identity is a maze or map that functions in a multitude of ways to guide and direct exchanges with one's social and material realities" (Cross 1991: 214; cited in Haslanger 2005). Haslanger notes that while some use the metaphor of a social script, analogizing identities to maps "might be preferable insofar as it need not be understood linguistically, and may

[6] Dembroff and Saint-Croix (2019) argue that labels are not needed for identities. However, they hold that identities do require having a conceptual representation of a group as an identity requires "an ability to recognize the group and to recognize one s own relevant similarity with it" (Dembroff and Saint-Croix 2019: 579). See Haslanger (2005) for arguments that Appiah's view of identity, and I take it Dembroff and Saint-Croix's as well, does not adequately account for people with white identities who do not take themselves to perform actions using a linguistic or conceptual representation that involves acting "as a White person" (2005: 277).

involve a 'map' of one's own body" (Haslanger 2005: 283). The relevant social and material realities include the distribution of various resources (economic, political, epistemic, …), the way social space is structured (e.g., in terms of bathrooms, grocery stores, elevators), and so on. Agents navigate these through combinations of norms, dispositions to acts, ways one conceptualizes one's body, history, actions, and so on. Alcoff takes identities, which she says are "grounded in social locations" (2006: 9) to have "causal determinacy over our epistemic and political orientations to the world" (2006: 90).

Using the notion of a social map, Haslanger and Jenkins offer views of racial and gender identity, respectively, as follows:

> "One has an X racial identity just in case their map is formed to guide someone marked as X through the social and material realties that are (in that context) characteristic of Xs as a group" (Haslanger 2005: 284)

> "S has a gender identity of X iff S's internal 'map' is formed to guide someone classed as a member of X gender through the social or material realities that are, in that context, characteristic of Xs as a class" (Jenkins 2016: 414)

On these views, having an identity is understood in terms of maps that are tied to a social group or class. This condition is important, I take it, as racial and gender identities are meant to be social, shareable, and not wholly private or individualistic.[7] Social classes/groups can have multiple members. They are caused to be and constituted by diverse social factors, not just individualist features or feelings, beliefs, or representations. For instance, as Jenkins argues, there is an objective element to identities as "there must be some genuine [perhaps imperfect] correspondence between the norms people take to be relevant to themselves and the norms associated with the relevant gender class in at least some context" (2016: 412). By tying identities to social locations or classes/groups, identities meet these constraints.

Since, on this view, identities always involve a relation to a social class/group, without the existence of the requisite relata, there will be no possibility of having an identity that involves a map formed to guide one based on the social or material realities of the (non-existent!) social class/group. If a map view of identity is adopted and we are aiming to determine whether there can be unlabeled identities, we face two related questions: Can there be unlabeled social groups/classes? And, if so, can there be unlabeled identities? I take it that the answer to both is yes.

[7] See Jenkins (2018) for an argument that relying on gender as class (or some other sort of gender roles that are not identical to gender identities) is also needed to avoid circularity.

Groups like Haslanger's children of widows, as well as fine-grained intersectional groups for which we do not have compact labels, are examples of groups that seemingly exist without labels. Haslanger (2005) and Jenkins (2018) both consider cases of identities that do not appear to have come into existence with the creation of a label. Haslanger (2005) argues that multi-racial identities involve having an aggregate or "substantially fragmented" map. These maps guide people who have them in ways that cohere with someone marked as being in one race in some contexts or on some dimensions and in ways that cohere with someone marked as of another race in other contexts or on other dimensions. Jenkins argues that a genderfluid identity can be understood in terms of, e.g., a map that sometimes guides a person in ways that are characteristic of those classed as a woman and at other times guides them in ways characteristic of those classed as a man (2018: 735-6). While Haslanger and Jenkins do not discuss a lack of terminology and we now have labels like *genderfluid* and *multiracial* it is compatible with their views that one could have a fragmented, aggregate, or fluid map without there being a label for a class or group of people the map is meant to guide.[8]

To put the argument more succinctly, groups, including those that are unlabeled, position people in ways that involve systematic applications of norms, constraints, enablements, and so on.[9] People positioned in these ways will plausibly come to internalize ways of moving through social reality that are sensitive to these norms, constraints, and so on. That is, they will plausibly come to have an identity tied to an unlabeled group even though they do not (and cannot!) self-label.

The ways one is guided by an identity might be opaque even to the person themself. This might even be the standard case.[10] Later, a person might interrogate their behavior or thought patterns—perhaps with the help of a therapist or close friend—and come to better understand how their identity was shaping the ways they felt, what options seemed available to them, how they interacted with others, and so on. This reveals a limitation with the map analogy, since with an actual map, one looks at it, interacts with it, and consciously uses it to guide their movement. In contrast, on the map view of identity, a gender map (or other identity map) can be internalized so that it guides a person without requiring them to have conscious access to it. It might function a bit more like a map

[8] While the account of an aggregate or fragmented map might be the right way to analyze some people's multiracial identities, others with multiracial identities might feel that there are contexts in which no racial norms seem to apply to them. Similarly, Jenkins argues that an agender identity involves not having a map that functions to guide one in ways that are characteristic of any gender class (2018: 736).

[9] Some take social categories (i.e., social groups or classes) to require representations. For instance, Mallon argues that representations of categories are a crucial part of the mechanism by which" categories are constructed (2016: 8). I do not take this to be a requirement on category/group/class construction.

[10] See Flores and Camp (forth.) for arguments that identities can be centered in one's ways of thinking about others and oneself without *beliefs* about them.

in a self-driving car—guiding one through patterns of action, conceptualization, etc. in ways that do not involve one thinking "these are the sorts of actions people with identity *I* or in group *G* (ought to) take".

Before moving on, I want to note that I do not mean to imply that labels can have no effects on people's identities. For instance, I do not endorse the following as a complete picture of identity label introduction.

> Stage 1: There are some people with identity, *I*, for which there is no label in a language *L*.
>
> Stage 2: A label, *N*, is introduced into the lexicon of *L* allowing people to more easily share that they are *I* and their *I*-related experiences.

On this sort of picture, the process of introducing a label *only* changes our linguistic (and perhaps conceptual and epistemic) resources. This is far too simplistic. Introducing labels can have metaphysical effects on identities as well. That does not require that there was no identity to begin with. (For there to be something to change, there must have been something there to change!) Rather, it involves the claim that labels can change identities. For instance, norms might be felt to apply more strongly and might be more strictly enforced. Hacking's notion of looping effects suggests one route by which this might occur (1995; 1999). In coming to be labeled or to self-label as an *N* the map that guides those who have identity *I* might be transformed. Appiah states that "labels operate to mold what we may call identification, the process through which individuals shape their projects—including their plans for their own lives and their conceptions of the good life—by reference to available labels, available identities" (2005: 66). While Appiah takes labeling to be a requirement for identity and I do not, I agree that labels can shape one's conceptions of oneself, the narratives one constructs, and so on.[11] To construe this within the map analogy, one might note that if a certain part of a mountain is labeled as the highest peak this might change the way people interact with the physical environment. For instance, hikers might seek a direct route to this point, thereby creating a path that will eventually come to be marked on subsequent maps. Someone labeling themself as an *N* might shape how they interact in the social world, how others interact with them, and so on.[12]

I have argued that there can be unlabeled identities by relying on the notion of identities as involving internalized maps. Given that there can be unlabeled identities, the question about what

[11] There is also evidence that the way one labels can affect how one conceptualizes oneself. For instance, Walton and Banaji (2004) found that when people were required to report preferences with noun constructions ("I am a coffee-drinker") rather than a verb phrase ("I drink coffee a lot") they took these preferences to be stronger, more stable, and more resilient. This suggests that self-labeling with nouns can shape self-perception.

[12] Thanks to Katharine Jenkins for suggesting this way of extending the analogy.

linguistic strategy ought to be adopted—to label or not to label—looms.[13] In the next two sections I consider the effects that can come from labeling and failing to label. I frame the discussion largely in terms of the harms that come with failures to label and the harms that come with labeling, although I recognize that several of the features cannot be neatly categorized as generally positive or negative. I return to considering the particularities of identities, power, and the value-laden effects of labels in *IV*.

II. Effects of Failures to Label

Failures to label mask the existence of identities and their corresponding social groups. It will emerge that some of the effects of labeling lacunae are general, while others vary with differences in power. In the case of dominant identities, masking serves to promote the interests of the group and to make their hegemony harder to see, theorize, and challenge. In the case of marginalized identities, masking is a form of hermeneutical injustice. It works to maintain marginalization and impedes forging bonds of solidarity. In the next three subsections I discuss general effects of the lack of labels and more specific effects of failing to label dominant and marginalized identities.

A. General Effects of a Lack of Labels

Shared terminology allows us to more easily collaborate and coordinate. Examples of this abound in every area of life. We use shared terminology in efforts to address climate change, to apply theories of differential equations to new domains, to coordinate to develop new technology, to differentiate and distinguish between styles of music, and to collaborate with a co-parent to raise children. Every field involves some terminology that allows for researchers, students, and others to more concisely express ideas that are then used in further theorizing and applications.

Shared terminology also allows us to offer explanations and justify or criticize the existence of descriptive and normative regularities. Nouns, one labeling device, have been argued to promote certain sorts of kind-based reasoning and explanation (Bigler and Liben 2007; Gelman 2003). Failing

[13] Some (but not all) of the arguments I give in the remainder of the paper could be used to support a project that involves both a linguistic introduction and a metaphysical creation component. For instance, the creation of an identity and the introduction of a label for it might allow for people to come to have more authentic identities, which would be beneficial. I won't consider this variant further, but it is open to someone who disagrees with the metaphysical arguments given in this section to pursue some of the same strategies I do below.

to have labels for social identities—either dominant or marginalized—can inhibit our ability to offer easily digestible explanations.

A lack of a compact label can impede explanation and prediction. For instance, consider a case in which A is asking for an explanation for why someone behaved in a particular way and B offers one of a variety of potential responses.

1. A: Why did Bradford do that?
   B: Because he's gay / Black / cis / white / a tall funny linguist.

While all of B's answers sound felicitous in some contexts, some of them seem suitable (as in, conversational participants would accept them as answering the question) in many contexts, while others are much more limited. Consider explanations in terms of the first two features. These could involve an appeal to stereotypes. For instance, one can easily imagine that B is homophobic or racist in appealing to Bradford being gay or Black as an explanation. However, taking being Black or gay to be explanatory need not involve derogation. For instance, suppose A is asking why Bradford put his hands high on the steering wheel while slowly and calmly asking the police officer who pulled him over whether he could reach into the glovebox for his registration. B responds: because he's Black. It is not just stereotypes about purportedly intrinsic biological features to which social explanations can appeal. Rather, social structure could also underpin why someone being Black, trans, and so on figure in explanations (Haslanger 2016; Ritchie 2019; Vasilyeva and Lombrozo 2020). This makes social identities potentially explanatory in a wide range of contexts. It is more difficult to imagine a context in which a longer expression like "a tall funny linguist" would be explanatory.[14] In many contexts, B answering in that way would not be taken to be a suitable response to A's question. Whether there is a lexicalized label, either in the form of a noun or an adjective, makes a difference when offering explanations that are taken to be acceptable.

Expressive resources impact how well one can perform on tasks involving categorization and matching. For example, Lupyan and Zettersten (2021) found that shorter descriptions of a distinction (in terms of the number of words or number of clauses used in the description) were correlated with greater accuracy in drawing a distinction between two categories. Some of their stimuli involve Bongard problems and variants of them (Bongard 1967). Bongard problems involve twelve

[14] See Prasada and Dillingham (2006, 2009) and Haward et al. (2018) for empirical work showing that people take lexicalized expressions to be better explanations than complex noun phrases like 'tall funny linguist.'

geometrical representations in two sets of six. Participants are tasked with finding the difference between the images in the two sets; this is the solution to the problem.

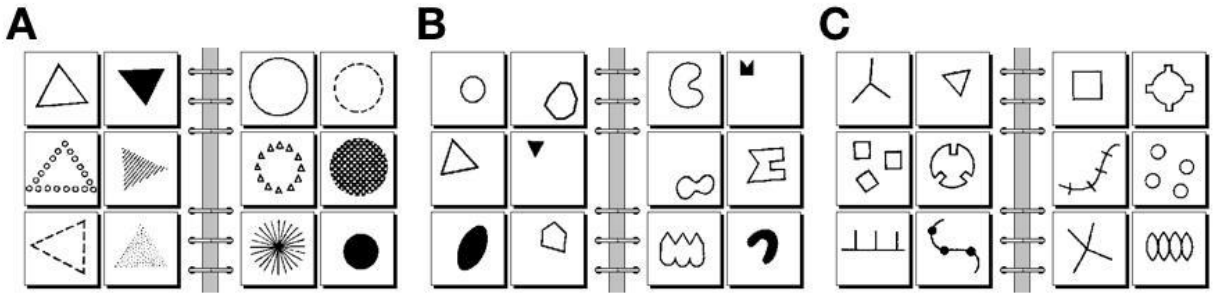

Figure 1. Three examples of Bongard Problems (image from Lupyan and Zettersten 2021)

For instance, in Example A in Figure 1, the set of figures on the left are all triangular but differ in size, borders, and so on. The figures on the right are all circular but also differ along other dimensions. This problem was easy for participants to solve. They also had lexicalized expressions to easily draw out the difference—*triangle* and *circle*. In contrast, distinctions that were more difficult to compactly describe in a participant's language were harder to solve. For instance, while the difference in the sets in Example B can be described as being convex or concave, not all participants are familiar with these labels and far fewer solved this problem. Overall, Lupyan and Zettersten found that lower linguistic complexity was significantly correlated with successfully solving a problem and with greater accuracy in categorizing new exemplars into a category. These findings are intriguing but they are consistent with either (i) compact labels facilitating categorization or (ii) certain categories being more natural for us to comprehend thereby both facilitating categorization judgments and explaining why the categories are named.

To differentiate between these Lupyan et al. (2018) tested several groups of participants with different access to formal education and language input (in the case of the latter, congenitally deaf Chinese children deprived of language input for much of their early lives). In a geometrical task requiring participants to find "the odd one out", they found that overt naming improved performance, that children with limited language input performed far worse than children with normal language input, and that verbal interference (repeating "a, b, c") selectively negatively affected participants responses on easy to name trials. These findings suggest that there is a connection between how easy

it is for us to verbalize a category distinction and how well we can draw the distinction and correctly classify new instances. Labels appear to help facilitate our categorizing practices—for good or for ill. Next let's consider the effects of unlabeled dominant identities.

B. Specific Effects of Unlabeled Dominant Identities

When dominant groups go unlabeled their hegemony can be masked. Categories that are unlabeled are less salient and will be taken to be less explanatory. There is evidence that even when there are labels, not *using* a label can have unjust effects and, of course, a label cannot be used if it does not exist at all. For example, returning to example 1 above, in some contexts explanations that appeal to being white or cis will be interpreted as non-answers or as uncooperative.[15] By failing to have or use labels, the dominant group can normalize their interests, norms, practices, and so on as features of a neutral "non-identity" individualistic stance (Du Bois 1924).

Sociolinguistic research has found that dominant groups in high school and university settings are often not labeled (Eckert and McConnell-Ginet 1995; Kiesling 2001). For instance, Eckert and McConnell-Ginet argue that "If the dominant category is not even labeled (and, as we noted earlier, in many schools it is not), then its distinctive interests are somewhat easier to ignore, its hegemonic control over social values and institutional norms more readily established. Two category labels in direct opposition reflect a live ongoing social struggle" (1995: 13). By failing to label themselves or be labeled by others the dominant group comes to enjoy an "unmarked default or 'normal 'status within their school" (McConnell-Ginet 2020: 71). Carrying this over to cases of unlabeled identities in a broader non-scholastic setting, McConnell-Ginet considers the introduction of the terms *cis*, *cisgender*, and *cishet*. She notes that "in many, many contexts, it still 'goes without saying' the people are cisgender" adding that cis people want to view themselves "as just 'normal' humans" (2020: 75).

Mills' notions of "an epistemology of ignorance" and "white ignorance" draw out a way those positioned as racially dominant can be ignorant of "the world they themselves have made" (1997: 18; see also Mills 2007). He goes on to say that "Part of what it means to be constructed as "white" …, part of what it requires to achieve Whiteness…, is a cognitive model that precludes self-transparency and genuine understanding of social realities" (Mills 1997: 18). Other dominant identities, including those that are unlabeled, can also be masked to both in-group and out-group members. The failure to have or use labels for dominant groups allows for dominant identities to be obscured even more.

[15] Of course, these identities can figure in explanations and in some contexts they will be considered highly explanatory (see, e.g., discussions and memes using the label *yt* on Twitter). The label *yt* is pronounced like *white*.

Masking and making dominant identities appear as the norm or as neutral non-identities, also makes individuals with non-dominant identities appear deviant. For example, consider accents. While these are not identities per se, they are clearly observable and can signal group membership or identity. Many take the Standard American accent (aka a broadcast English accent) to be an accent-less form of English. That is, they take it to involve no accent whatsoever and would not use a label like *Standard American accent*. On this conceptualization, other accents involve divergence from a norm or standard. They may be taken to involve mispronunciations, or to be funny, cute, or impossible to understand. What is taken to be standard is taken to be normal. And this normality is not just in terms of a (often false) view about statistical likelihood, but in terms of what is correct. On this view, it is not just that divergence from the norm is statistically odd, but that it is wrong; it is deviant in the morally-laden sense of the word.

C. Specific Effects of Unlabeled Marginalized Identities

The lack of an identity labels for marginalized groups can make it harder for members to make their experience and social situation intelligible, constituting a form of epistemic injustice. Let's begin by focusing on a broader failure in having conceptual or linguistic resources. In discussing nonbinary people, Dembroff argues that a conceptual lacuna "contributes to a hermeneutical injustice that arises from the failure to spread and charitably analyze the concepts and practices underlying nonbinary classifications" (2020: 2). As Fricker defines it, hermeneutical injustice is "the injustice of having some significant area of one's social experience obscured from collective understanding owing to a structural prejudice in the collective hermeneutical resource" (Fricker 2006: 100). A limitation in linguistic, conceptual, and epistemic resources affects what can be shared, discussed, and made intelligible (Dotson 2012; Pohlhaus 2011; Fricker 2006, 2007; Medina 2013).

Limitations in linguistic, conceptual, and epistemic resources can render an aspect of someone unintelligible to everyone or to just some people. Clearly, if in a discourse context or socio-linguistic community there is no label for an identity it will be difficult, if not impossible, to communicate or share one's identity with others. As Mason (2011) and Dotson (2012) argue, this does not show that understanding or knowledge is impossible, full stop. Rather, they argue, even if the resources of the dominant discourse are limited, there can be alternative resources that surpass those used in dominant groups, and which allow for one with access to them to know and render intelligible their identities and experiences. In some cases, though, both dominant resources and alternative non-dominant

hermeneutical resources can lack the tools to communicate about an identity. In this case, a person's identity will be difficult or impossible to interpret or understand for others and for themself.[16]

A lack of labels can also make organizing and fostering solidarity challenging. It will be more difficult to find commonalities and band together to work to mitigate oppression that comes from marginalization without linguistic resources for shared identities and shared forms of experience. Labels are one tool that facilitate, or perhaps better, serve as a precondition for organizing.

To write laws and advocate for policies to abate the marginalization and oppression faced by an identity group, lawmakers and politicians need to have access to linguistic resources to pick out the relevant group. For instance, in implementing affirmative action or instating reparations for past injustice conceptions of the groups that a law is aiming to serve are needed. As discussed in section *IIA* above, having a compact label, rather than a long description, is apt to mark out a category more clearly and figure in more digestible explanations. Linguistic resources of some sort are needed to advocate for new anti-oppression policy; labels help facilitate making these clear and compelling.

Finally, there can be metaphysical effects that come from a lack of resources to conceptualize and discuss identities. While I have argued that there can be unlabeled identities, the lack of resources to talk and think about them limits how people see others and how individuals conceptualize themselves. This could lead some to being seen as having or actually coming to have identities that are not authentic to themselves. For instance, Fricker suggests that "hermeneutical injustice can mean that someone is socially constituted as, and perhaps even caused to be, something they are not, and which it is against their interests to be seen to be … Hermeneutical injustice is an epistemic injustice with social constructive power" (2006: 107-8). On a similar note, Ásta states, "Sometimes there is no location in the current context that we identify with. And sometimes there are very few contexts with locations that we identify with. Our struggle then is to make it the case that there are more contexts in which there are locations that we can identify with" (2018: 123). There is a strong interpretation of both Fricker and Ásta on which they are saying that one cannot have an identity without it being labeled, conceptualized, or contextually recognized. Given my arguments in *I*, I do not adopt this strong view. However, I agree that linguistic or conceptual limitations can elicit false or unjust

[16] I do not take this to require that all thinking is done in language. Rather, it is to say that without shared linguistic or conceptual resources for an identity, representing it at all, even if it is an aspect of one s own identity, can be difficult or impossible. There are interesting issues regarding how discursive and iconic representational formats relate to identities, that I cannot address here but I hope to in future work. See Flores and Camp (forth.) for discussion of centering identities with either images or linguistic representations.

perceptions of who or what a person is[17] and, perhaps even more troubling, could shape people in ways that give rise to dysphoria and feeling inauthentic.

III. Effects of Labeling

Some of the preceding discussion makes clear what introducing and using a label is apt to do. Introducing labels for both powerful and marginalized groups can serve to unmask. In the case of the powerful, this can lead to seeing what might be viewed simply as "the norm" as in the interest of dominant groups. In the case of marginalized groups, this can serve to make identities more salient and shareable, allowing for greater solidarity and organizing power. Insofar as labels facilitate drawing category distinctions, explaining, and categorizing members into kinds, introducing labels where there were none will make these tasks easier.

Using labels can also have harmful effects. It has been argued that labels divide, lead to cognitive and normative distortions, and increase bias, overgeneralization, and essentialism. As I stated above, much of the discussion in *II* focused on the harms of labeling lacunae. Much of the discussion in this section is focused on the harms of labeling.

Labels have been argued to be divisive. Sometimes these arguments come from those with power, for instance by white people wanting to focus less on race or cis people not wanting to be called *cis* and pathologizing being trans. As McConnell-Ginet nicely puts it "Those who *can* resist being marked often do so. They decry labels as 'divisive' and affirm people's individual distinctiveness. Treat us as individuals, they say. Don't lump us all together. Don't generalize" (2020: 76, *emphasis original*). McConnell-Ginet's words here echo arguments that have been given against identity politics and that are used to deny the relevance of race, gender, sexual orientation, and other identities. There is, of course, also something compelling about these sorts of arguments—labels certainly can facilitate overgeneralizing.

In psychology there is a large body of research on the effects of groups and group membership on performance, preferences, and judgments. For instance, people tend to have preferences for in-group members (Tajfel et al. 1971). Using generics and labels increases in-group preferences and leads

[17] Some metaphysical views of social categories require a bit more nuance on this claim. For instance, Ásta (2018) would reject the idea that a person with adequate standing in a context can falsely take someone to be a member of a social category. On her account if one is taken to have some identity (i.e., taken to be a member of some social category) by someone with adequate standing, one really is a member of that category. She would, I believe, take the phrasing in terms of unjust categorization to be acceptable.

to positive information being generalized more to in-group members while negative information is generalized more to out-group members (Bigler et al. 1997; Baron and Dunham 2015; see Liberman et al. 2017 for an overview and additional citations). These effects have been found even in minimal groups like those based on matching shirts that are randomly assigned to participants. Introducing labels for identities that are more robust and not simply made-up in the setting of a psychology experiment is likely to have the same or stronger effects.

Research points in different directions as to whether labels are *required* for in-group and out-group effects (see Baron and Dunham, 2015; Baron et al. 2014 for evidence that labels are needed; see Jordan and Dunham 2020 for arguments that labels are not needed). Yet, regardless of the answer to this question, evidence shows that labels do augment in-group/out-group effects.

Labels have also been argued to promote psychological essentialism. Psychological essentialism is the view that we take some categories to be inductively rich, stable, and explanatory and to have an underlying unobservable essence that causes or otherwise realizes observable features (Gelman 2003; Newman and Knobe 2019; Neufeld 2022). While not all labels appear to promote essentialism (e.g., nouns for artifacts like *hammer* do not, or do not to the same degree), some research finds that nouns for human groups like racial and gender terms do augment essentialist thinking. For instance, consider the generics in 2 and 3.

2. Asians are good at math.
3. Women are better caregivers than men.

The generics in 2 and 3 express stereotypes and may lead to strong judgments that there is a biological essence of the social kinds Asians and women that causes their members to share these features (Rhodes et al. 2012; Cimpian and Markman 2011). Labels used outside of generics, as in 4, or in conjunction with images have also been argued to essentialize (e.g., Gelman and Heyman 1999; Gelman 2003; Carnaghi et al. 2008; Ritchie 2021b).

4. Sara is a female / a queer.

This has led some to argue that we ought to avoid social generics or even social labels more generally (Haslanger 2011; Leslie 2017; Wodak and Leslie, 2017; but see Saul 2017; Ritchie 2019, 2021a).

Yet, other research suggests that labels for social identities do not always promote essentialism. They may, for instance, be accepted in virtue of social and contextual features that are not related to individual kind members having a biological essence (Vasilyeva and Lombrozo 2020; Noyes et al. 2021). Moreover, while essentialism can have negative effects, it is not always correlated with increased prejudice and in some cases is even correlated with *reduced* prejudicial attitudes (Haslam et al. 2002;

Peretz-Lange 2021). Labels clearly have significant effects on social cognition, but what generalizations hold is a topic that is still being actively studied.

IV. Introducing Labels for the Unlabeled

We have seen that there are harmful effects of labeling lacunae. There are also harms that can come with labeling. There are myriad complexities involving the introduction of labels. No simple prescription can adequately navigate these. Here I consider introducing labels for an unlabeled dominant identity and an unlabeled marginalized identity and argue that there are cases in which both introductions can bring benefits. Moreover, I argue that harms may not result, and steps can be taken to mitigate their possible occurrence. Using both examples illustrates the differences power plays in label introduction and shows that benefits are not intrinsically tied to the social standing of an identity group. Overall, I conclude that appeals to identities and the use of identity labels can be positive. Labels cannot be simply and summarily dismissed as harmful.

The introduction of labels for both unlabeled dominant and marginalized identities has positive epistemic effects. We saw that one primary harm of failing to have or use labels for dominant identities is that the group's hegemony and the particularity of their interests are masked. By labeling a previously unlabeled dominant identity, it becomes clearer that this is not a neutral identity. The label can be used to describe and explain, perhaps revealing to us another form of hierarchical social structure. If all identities—including dominant identities—are labeled, this can also help to mitigate the view that there is something deviant about non-dominant identities. In the cases I am focused on, labels are not creating new identities. Rather, introducing labels allows for the existence of identities and the power and normative dynamics they involve to be brought out into the open. This, in turn, could change the dynamics of power and identities themselves (perhaps through looping effects (Hacking 1995, 1999)). Yet, even in these cases, an identity is not being created ex nihilo with the introduction of a label.

Introducing a label for a previously unlabeled marginalized identity, perhaps as a part of a broader "conceptual revolution" (Langton 2010: 460), works to alleviate a hermeneutical injustice. Linguistic introduction and conceptual overhauling are strategies used to remedy the lack of resources available to make sense of and communicate identities. As Pohlhaus (2012: 719) explains:

> when there is a tension between the world of experience and the resources that we use to make sense of our experiences, for example when the proper

> language for describing an experience appears to be missing, or when our current concepts fail to track recurring patterns, we recalibrate our epistemic resources and/or create new ones until the tension between our resources and the experienced world is alleviated.

Allowing people to better understand, describe, and explain their identities and experiences are all positive effects of introducing a label for a previously unlabeled marginalized identity.

Expanding linguistic resources has epistemic benefits for anyone with access to the resources. The other benefits and harms of using labels vary depending on group status and whether the label is used for self/in-group members or for others/out-group members. While I argue that both introductions can bring important benefits that are not always outweighed by harms, it is important to note once again that I am not offering a simple prescription to always increase labels or labeling. There are clear cases in which identities are appealed to in harmful ways. I do not take this to be up for debate. Even the most committed proponent of identity politics or person who takes Black studies, feminist philosophy, queer studies, and so on to be vital academic projects will openly accept that there are some instances in which identities and identity labels are used in harmful ways. For instance, a use of a label might be intended to demean. I take it whole classes of expressions, like slurs, ought not be used (except perhaps in cases of appropriation).

The cases I consider here do not involve introducing new *derogatory* terms for out-group members. Often, but not always, derogatory terms are used to label groups that are already labeled. For instance, Tirrell (2012) discusses the introduction of new derogatory terms for Tutsi by Hutu preceding and during the Rwandan genocide. She argues that these were used amongst Hutu (e.g., by out-group members) to emphasize differences between groups that were already labeled in other ways, to dehumanize, and ultimately as one way to put in place norms that made murder acceptable. As I will argue, the epistemic access to the existence of an identity and whose interest it will be in to introduce new linguistic tools differ starkly between the cases of introducing a new label for an already labeled identity and introducing a label for a previously unlabeled identity. I make the case that introducing labels for the unlabeled can be beneficial by considering first dominant and then marginalized identity labels.

Consider an introduction of a label, *D*, for a previously unlabeled dominant identity. (I will work with the abstract *D* but it may be useful to think of *cis* as an example of a recently introduced label for a previously unlabeled dominant identity.) When out-group members begin to label the dominant identity using *D*, those with non-dominant identities can better describe and explain the

features just discussed. The label can help in consciousness raising and in remedying epistemic occlusion. These are all benefits.

These beneficial effects are not limited to uses of *D* by out-group members. In-group/self-labeling with *D* can also involve recognizing, rather than taking for granted, that one has a dominant group identity. In this way, self/in-group labeling can also be positive.

Of course, not all instances of self/in-group labeling by people with a dominant identity are positive. Even well-intentioned people with dominant identities, can use self/in-group labeling in ways that continue to center powerful groups. For instance, think of people who frequently describe themselves as having white privilege. This can involve an apologetic stance or a feeling of self-hatred. For instance, in an anecdote I recently heard, someone said they hated their white skin to such a degree that they claimed to want to tear it off. While recognizing privilege is a valuable step in undermining oppression, focusing on the privilege of those in power is not likely to have significant remedial effects on power imbalances, and may reinforce the power dynamics already in play.

There is a second sort of self/in-group labeling in the case of dominant identities that brings with it the potential for serious harms. As we saw, using labels can heighten inter-group effects, including increasing preferences for in-group members and judging that out-group members are more apt to do bad deeds than in-group members. If members with a dominant identity begin to use *D* to talk about themselves, this could lead to strengthened preference for other Ds. Uses like these are manifest in the language of dominant identity group movements—like white power movements. While clearly the language used in these movements is not the only factor that is relevant when considering the increased prejudice and violence they bring, it is one component.

The cases just considered show that in-group uses of labels by dominant group members have a range of effects. To show how these can range from positive to extremely negative, let's consider three members of a dominant identity group D who use *D* to label themselves and their in-group. One person strongly identifies as D, one is apologetic about their membership in D, and one recognizes their membership in D without strong identification or apology. Since the first strongly identifies as a D they are most likely to use *D* in ways that will promote further solidarity—*we Ds are the ones actually getting the short end of the stick!*—thereby continuing to advantage and bolster the extant dominance hierarchy. The apologetic figure will continue to center dominant identities in discourse in ways that sustain some features of hierarchies (DiAngelo 2018). The third might reflectively engage in using *D* in positive ways—recognizing and expressing that there are structural forms of oppression,

not denying their identity, and actively working to promote change. These cases suggest that for dominant group members, stronger self-identification with a label is apt to elicit more negative effects.

Next, let's consider who is likely to introduce a label for D. By being unlabeled, the dominant group continues to mask their hegemony and retain power. Insofar as it is in the interest of the dominant group to retain their dominance in this way, a label is likely to be introduced and used by non-group members first. In such cases, out-group uses are likely to be positive; they involve allowing for explanation, consciousness raising, and organizing that would be difficult or impossible without shared resources.

It is also relatively rare for dominant group members to self-label or use in-group labels even when labels exist (e.g., recall discussion of *cis* and *heterosexual*). So, it is possible there will be almost no in-group/self-labeling. If there is in-group/self-labeling some of it may also be genuinely positive (aimed at recognition and change). Building on this case, uses of *D* that are most likely to have harmful effects are those that involve reaction and a feeling of the loss of power and these uses may not develop at all. This shows one sort of case in which introducing a label for an unlabeled dominant group is apt to bring important benefits and suggests that these will not always be outweighed, given interests in not self-labeling.

Other cases are different. Dominant groups can also have interests in making their dominance manifest, particularly if they aim to solidify it further through an active movement to repress others in society. In this case the introduction of a label is unlikely to bring greater benefits than harms. In fact, it might be a tool primarily used to oppress.

Both sorts of cases are possibilities. My aim here is to show that given the interests dominant group members often have in not having or using labels for their identities (as discussed in *IIB*), there is a class of cases in which labels will be introduced by out-group members with marginalized identities. These bring epistemic and social and political benefits. In-group/self-labeling may only rarely occur, and it too could be positive. There are good reasons to think that introducing a label *D* can bring real benefits.

Next, consider an introduction of a label, *M*, for a previously unlabeled marginalized identity group. (I will work with the abstract *M* in this case, but to make the case more concrete one might think of *non-binary* as an example of a recently introduced label for a previously unlabeled marginalized identity.) In-group uses of the label might support greater self-understanding and diminish feelings of dysphoria. It can also facilitate organizing and help to promote unity and cohesion. For instance, in considering the way in which labels are used in class struggle, Iris Marion Young states "One can and

many do say, 'I am a worker,' as a badge of pride and identity….when this happens … one has formed a *group* with other workers with whom one has established self-conscious bonds of solidarity" (1994: 727). In addition to being organizing tools, in-group uses of identity labels might also, as Young suggests, change or create new groups, which may become hubs for increased political power (see also, Ritchie 2021c).

Out-group uses of *M* could also be in the service of greater recognition of M in ways that are positive. But not all out-group uses will be beneficial. They could involve mere symbolism, without doing anything to deconstruct hierarchies (Táíwò 2022). They might also involve labeling in ways that increase stereotypes, out-group bias, and essentialism. These sorts of harms are connected to the extent to which an out-group label is used in a way that "otherizes" members of the group. The greater degree to which the use of an identity label seems to be saying that *they* have some feature in a way that contrasts with features *we* have, the higher the likelihood that it will have negative effects. The pattern here is the inverse of that which I argued holds for dominant self-labeling. In that case, the more a person self-identifies with the identity group, the higher the likelihood of negative effects of in-group labels, like *D*. Here, the more one represents a marginalized out-group as distinct and other, the more likely it is for uses of *M* to have detrimental social and psychological effects.

Given interests and epistemic access, new labels for unlabeled marginalized identities are likely to be introduced by those who have the marginalized identity. It is people with the identity who recognize that there is a, as Pohlhaus put it, "tension between our resources and the experienced world." Individuals with marginalized identities are also the group with the strongest interest in having tools to better conceptualize and discuss their identities. For these reasons, in-group usage and self-labeling are likely to precede out-group/other labeling. Further, as Mason (2011) and Dotson (2012) argue, new resources might be introduced as alternative resources or resources that are more widely available in dominant discourse. If the introduction is only in an alternative discourse that is shaped by those with the identity, it is likely to involve in-group and self-labeling as the discourse and resources are being shaped, at least in part, by those with the once unlabeled identity. These uses bring with them important positive effects—epistemic, social, psychological. Harms may not result, and steps can be taken to mitigate them. For instance, labels might be used by those within particular communities without aiming to introduce them into the wider dominant discourse. Strategies could be developed to work to avoid the appropriation of conceptual resources by powerful groups (Davis, 2018) or elite capture (Táíwò 2020, 2022).

This shows that there is a class of cases in which the introduction of labels for marginalized identities is apt to bring benefits. Harms may come about as well, but the epistemic, political, social, and other benefits are real factors speaking in favor of augmenting linguistic resources.

V. Concluding Remarks

I've argued that uses of labels for both marginalized and dominant identities can have positive and negative effects. By considering the interplay of epistemic access and interests tied to power, I have argued that there are classes of cases in which introducing labels for unlabeled identities brings important (epistemic, social, political, cognitive, …) benefits and that given interests and alternative strategies, harms can be mitigated. The social world and the way we think and talk about it is extraordinarily complex. I do not mean to neglect this. Rather, my aim here was to push back on the anti-identity and anti-identity language strand in some popular and academic discourse. Identity labels can bring real benefits that are not always outweighed by harms. A more nuanced position recognizing that even adding more identity labels can bring real benefits is needed.

I close by considering three final worries. First, someone might argue that while perhaps not all labels do more harm than good, the potential for harm from a label is astronomical, suggesting that we ought to reduce labeling with extant terminology and avoid introducing any new labels. To harken back once more to the tool metaphor, the argument could be framed in terms of taking identity and other social labels to be more like guns (or even nuclear weapons) than hammers. The claim is then that labels have the potential to be so detrimental that they ought not be available to people at all. I have argued against this extremely strong view by considering the epistemic, psychological, social, and political benefits of labels and labeling.[18]

Second, one might object that I have failed to consider the longer-term effects of introducing labels for unlabeled identities. And, the argument might go, once a broader time scale is considered it will be found that labels are more harmful than beneficial. As a generalization, this is an extremely strong thesis. The ways that racial, sexual orientation, and gender labels have figured in successful struggles for rights and protections provide reason to think it is too strong and that it should be rejected. However, it might still be pressed, and rightly so, that there are longer-term harms that need to be considered.

[18] I thank an anonymous referee for suggesting this concern and way of extending the tool analogy.

To make the case in a vivid way, let's focus on the introduction of a label for a previously unlabeled marginalized identity *M*. Suppose that while there are some initial positive effects from in-group and self-labeling with *M*, after some time the label is used by those with dominant identities to stereotype people who are M, promote bias against Ms, and ultimately to carry out atrocities that include murdering many people who are M because they are M. Clearly this is a heinous scenario. But what we should conclude about the introduction of the label is, I think, not obvious.

The label plays a role in the story, but it is not the only factor. No one would argue that the introduction of the label was *the* cause of genocide.[19] The power dynamics that existed below the surface of the linguistically available resources were there before. In fact, the relationship between the introduction of the label and the genocide is plausibly highly sensitive to many changes in background conditions, making it a less appealing causal explanation (Lewis 1986; Woodward 2006, 2010, 2021). But, given what I have argued here, the introduction of the label did contribute to bringing these to the surface.

Would it have been better if the label were never introduced? Certainly it would be better if the genocide had not occurred, but precisely what the role of a label for Ms played in the story is not easy to pin down. There are possible scenarios in which a label was introduced, it led to benefits like Ms better understanding their identities and banding together to improve their standing in society, and it did not lead to the atrocities in this imagined scenario. I take cases like these to illustrate once again the intricate and complex connections between language, identity, and power that have and ought to be further explored, but which a single paper cannot fully address. My claim is quite humble—there are cases in which introducing labels for unlabeled identities is beneficial. These ought to be recognized and reveal that any general claim that identity labels are bad and should be avoided is too strong.

Finally, one might argue that it would be better for there to be no identity labels at all. This argument could be made in a more or less idealized way. As a point about what linguistic resources we ought to have given the social world as we find it, I take this line of thought to be unconvincing. Here's an argument to show why. I take it there are identities (i.e., they exist). Further, it is better to have linguistic resources to talk about entities (very broadly construed) that exist than failing to have such resources. So, it is better for us to have linguistic resources to talk about identities (i.e., identity

[19] See Tirrell (2012) for discussion of the ways language is one piece of how the Rwandan genocide came to be. She emphasizes, citing Semelin, that the role of language and propaganda should not be overemphasized as "there is nothing to prove that this, on its own, leads to the unleashing of a massacre" (Semelin 2007: 199).

labels) than not, given the actual ontology of social reality.[20] It may be better to live in a world without any identities (although I do not really know what such a world would be like). If so, it *might* be better to have no identity labels. One might argue that in such a world labels would not be *tracking* anything in reality and would only *create* divisions.[21] Here I did not consider this (fictional) scenario, but rather asked about the benefits and harms of introducing labels in the social world as we presently find it. If the ontology of the social world were radically different, the effects of identity labels and whether they ought to be in a language at all might also differ.

My aim here was to investigate a small part of the complex dynamics of language, power, knowledge, and identity by considering the introduction of identity labels. I argued that the introduction and usage of labels for existent but previously unlabeled dominant and marginalized identities brings (epistemic, social, political, cognitive, …) benefits. Even given the complexity and particularities found across situations and contexts, there are reasons in favor of augmenting a linguistic toolkit to include further identity labels.

Acknowledgements

Previous versions of this paper were presented at the University of California, Irvine Identities and Epistemic Injustice Workshop, 2021 Social Metaphysics Conference, the Vermont Mind, Language, and Social Change Workshop, the Pacific APA, the Conceptual Engineering Online Seminar, the Words Workshop, and at colloquia at the University of Vienna (in the Vienna Colloquia in Analytic Philosophy series), the University of Illinois—Urbana Champaign, and USC. I thank these audiences for their helpful feedback. In particular, I thank Lauren Abruzzo, Anna Boncompagni, Jeremy Evans, Carolina Flores, Katharine Jenkins, Zoë Johnson King, Alex Madva, Rebecca Mason, Eliot Michaelson, Bernhard Nickel, Daniel Wodak, 2 editors and an anonymous referee for suggestions and comments that significantly improved the paper.

Works Cited

Alcoff, L. M. (2006). *Visible Identities: Race, Gender, and the Self*. New York, NY: Oxford University Press.

Appiah, K. A. (1996). Race, Culture, Identity: Misunderstood Connections. *The Tanner Lectures on Human Values* 17, 51-136.

—— (2005). *The Ethics of Identity*. Princeton University Press.

[20] This is not to say that it is better to have derogatory labels, of course. But it is a very strong and controversial view that all identity labels are like slurs, for instance.

[21] Although there are reasons to think that there are benefits to labeling in promoting agency, see Flores and Camp (ms).

Ásta. (2018). *Categories We Live by: The Construction of Sex, Gender, Race, and Other Social Categories By*. Oxford University Press.

Baron, A. S., & Dunham, Y. (2015). Representing 'Us' and 'Them': Building blocks of intergroup cognition. *Journal of Cognition and Development*, 16(5), 780–801.

Baron, A. S., Dunham, Y., Banaji, M., & Carey, S. (2014). Constraints on the acquisition of social category concepts. *Journal of Cognition and Development*, 15(2), 238–268.

Bigler R.S., Jones, L.C., Lobliner, D.B. (1997). Social categorization and the formation intergroup attitudes in children. *Child Development* 68(3), 530–543.

Bigler, R. S. and Liben, L. S. (2007) Developmental Intergroup Theory: Explaining and Reducing Chidren's Social Stereotyping and Prejudice. *Current Directions in Psychological Science*. 16 (3), 162-166.

Bongard, M. M. (1967). *Pattern Recognition*. Hayden Book Co., Spartan Books.

Brooks, M. (1997). *Blurring the Edges*. Capitol Records

Burgess, A., Cappelen, H. and D. Plunkett. (2020). *Conceptual Engineering and Conceptual Ethics*. Oxford: Oxford University Press.

Burgess, A. and D. Plunkett. (2013a). Conceptual Ethics I. *Philosophy Compass*, 8(12), 1091– 1101.

—— (2013b). Conceptual Ethics II. *Philosophy Compass*, 8(12), 1102– 1110.

—— (2020). On the relation between conceptual engineering and conceptual ethics. *Ratio* 33(4), 281-294.

Cantalamessa, E.A. (2019). Disability studies, conceptual engineering, and conceptual activism. *Inquiry: An Interdisciplinary Journal of Philosophy*, 1–31.

Carnaghi, A., Maass, A., Gresta, S., Bianchi, M., Cadinu, M., & Arcuri, L. (2008). Nomina sunt omina: on the inductive potential of nouns and adjectives in person perception. *Journal of personality and social psychology*, *94*(5), 839–859. https://doi.org/10.1037/0022-3514.94.5.839

Cappelen, H. (2018). *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.

Cikara, M., Bruneau, E.G., & Saxe, R.R. (2011) Us and them: intergroup failures of empathy. *Current Directions in Psychological Science* 20(3), 149–153.

Cimpian, A., & Markman, E. M. (2011). The generic/non-generic distinction influences how children interpret new information about social others. *Child Development*, 82, 471–492.

Cross Jr., William E. (1991). *Shades of Black: Diversity in African-American Identity*. Philadelphia: Temple University Press.

Davis, E. (2018). On Epistemic Appropriation. *Ethics* 128 (4), 702-727.

Dembroff, R. (2020). Beyond Binary: Genderqueer as Critical Gender Kind. *Philosophers' Imprint* 20 (9), 1-23.

Dembroff, R. & Saint-Croix, C. (2019). 'Yep, I'm Gay': Understanding Agential Identity. *Ergo: An Open Access Journal of Philosophy* 6, 571-599.

DiAngelo, R. (2018). *White fragility: Why it's so hard for white people to talk about racism.* Boston, MA: Beacon Press.

Dotson, K. (2012). A Cautionary Tale: On Limiting Epistemic Oppression. *Frontiers: A Journal of Women Studies* 33 (1), 24-47.

Du Bois,W.E.B. (1924). *The Gift of Black Folk: the Negroes in the Making of America*. New York: Washington Square Press.

Eckert, P. & McConnell-Ginet, S. (1995). Constructing meaning, constructing selves: Snapshots of language, gender and class from Belton High. In Kira Hall and Mary Buchholtz (eds.) *Gender articulated: Arrangements of language and the socially constructed self*. London and New York: Routledge, 469-507.

Flores, C. & Camp, E. (forth.) "That's all you really are": Centering social identities without essentialist beliefs. In Haslanger, S., et al. (ed.) *Mind, Language, and Social Hierarchy*. Oxford University Press.

—— (ms). Identity labels as interpretive frames for building agency.

Fricker, M. (2006). Powerlessness and social interpretation. *Episteme* 3 (1-2), 96-108.

—— (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.

Gelman, S. A. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought*: Oxford University Press, USA.

Gelman, S. A., & Heyman, G.D. (1999). Carrot-Eaters and Creature-Believers: The Effects of Lexicalization on Children's Inferences About Social Categories. *Psychological Science*, 10(6), 489–493.

Hacking, I. (1995). The looping effects of human kinds. In *Causal Cognition: A Multidisciplinary Approach*, ed. Dan Sperber, David Premack, & Ann J. Premack. Oxford: Claredon Press.

—— (1999). *The Social Construction of What?*. Harvard University Press.

Haslam, N., Rothschild, L, & Ernst, D. (2002). Are essentialist beliefs associated with prejudice? *British Journal of Social Psychology*, 21, 87-100.

Haslanger, S. (2003). Social Construction: The "Debunking" Project. In F. Schmitt (ed.), *Socializing Metaphysics*. Lanham: Rowman & Littlefield, 301–325.

—— (2005). You mixed? Racial identity without racial biology. In S. Haslanger & C. Witt (eds.), *Adoption Matters: Philosophical and Feminist Essays*. Ithaca: Cornell University Press.

—— (2011). Ideology, Generics, and Common Ground. In C. Witt (ed.), *Feminist Metaphysics*. Springer Verlag, 179--207.

—— (2016). What is a (social) structural explanation? *Philosophical Studies* 173 (1), 113-130.

Haward, P, Wagner, L., Carey, S. & Prasada, S. (2018). The development of principled connections and kind representations. *Cognition* 176, 255-268.

Isaac, M. G., Koch, S., & Nefdt, R. (2022). Conceptual engineering: A road map to practice. *Philosophy Compass*, 17 (10), e12879, 1–15. doi: 10.1111/phc3.12879.

Isaac, M. G. (2021a). What Should Conceptual Engineering Be All About? *Philosophia Philosophical Quarterly of Israel* (5), 1–13.

—— (2021b). Broad-Spectrum Conceptual Engineering. *Ratio: An International Journal for Analytic Philosophy* 34 (4), 286–302.

Jenkins, K. (2016). Amelioration and Inclusion: Gender Identity and the Concept of Woman. *Ethics* 126 (2), 394-421.

—— (2018). Toward an Account of Gender Identity. *Ergo: An Open Access Journal of Philosophy* 5.

Jordan, A. & Dunham, Y. (2021). Are category labels primary Children use similarities to reason about social groups. *Developmental Science* 24 (1), 1-13.

Kiesling, S. (2001). Stances of Whiteness and Hegemony in Fraternity Men's Discourse. *Journal of Linguistic Anthropology* 11(1), 101-115.

Khalidi, Muhammad Ali (2013). Three Kinds of Social Kinds. *Philosophy and Phenomenological Research* 90 (1), 96-112.

Langton, R. (2010). Review of Epistemic Injustice: Power and the Ethics of Knowing By Miranda Fricker. *Hypatia* 24(2), 459-464.

Leslie, Sarah-Jane (2017). The Original Sin of Cognition: Fear Prejudice, and Generalization. *Journal of Philosophy* 114 (8), 393-421.

Lewis, David (1986). Postscripts to 'Causation'. In *Philosophical Papers Vol. II*. Oxford University Press.

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). The Origins of Social Categorization. *Trends in Cognitive Sciences*, 21(7), 556–568.

Lupyan, G., Wendorf, A., Rojas-Berscia, L. M., & Paul, J. (2018). Core Knowledge or language-augmented cognition? The case of geometric reasoning. *Proceedings of the 12th International Conference on the Evolution of Language (Evolang12).*

Lupyan, G., & Zettersten, M. (2021). Does vocabulary help structure the mind? In Sera, M., & Koenig, M. (Eds.). *Minnesota Symposia on Child Psychology: Human Communication: Origins, Mechanisms, and Functions, Volume 40*. Hoboken, NJ: John Wiley & Sons, 160-199.

Mallon, R. (2016). *The Construction of Human Kinds*. Oxford: Oxford University Press.

Markman, E. (1989). *Categorization and Naming in Children: Problems of Induction.* Cambridge, MA: MIT Press.

Mason, R. (2011). Two Kinds of Unknowing. *Hypatia* 26 (2), 294-307.

McConnell-Ginet, S. (2020). *Words Matter: Meaning and Power*. Cambridge: Cambridge University Press.

Medina, J. (2012). *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations.* Oxford University Press.

Mills, C. (1997). *The Racial Contract.* Ithaca, NY: Cornell University Press.

—— (2007). White Ignorance. In Shannon Sullivan & Nancy Tuana (eds.), *Race and Epistemologies of Ignorance*. Albany, NY: State Univ of New York Press, 11-38.

Nado, J. (2019). Conceptual engineering, truth, and efficacy. *Synthese*, 1–21.

Neufeld, Eleonore (2022). Psychological Essentialism and the Structure of Concepts. *Philosophy Compass* 17 (5).

Newman, G. & Knobe, J. (2019). The essence of essentialism. *Mind and Language* 34 (5), 585-605.

Noyes, A., Dunham, Y., Keil, F. C., & Ritchie, K. (2021). Evidence for multiple sources of inductive potential: Occupations and their relations to social institutions. *Cognitive Psychology* 130. 130, 101422. https://doi.org/10.1016/j.cogpsych.2021.101422

Peretz-Lange, R. (2021). Why does social essentialism sometimes promote, and other times mitigate, prejudice development? A causal discounting perspective. *Cognitive Development,* 59, 101085.

Pohlhaus, G. (2012). Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance. *Hypatia* 27 (4), 715-735.

Prasada, S., & Dillingham, E. (2006). Principled and statistical connections in common sense

conception. *Cognition*, 99 (1).

Prasada, S., & Dillingham, E. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, 33.

Rhodes, M., Leslie, S. J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, 109(34), 13526-13531.

Ritchie, K. (2019). Should We Use Racial and Gender Generics? *Thought: A Journal of Philosophy* 8 (1):33-41.

——(2021a). Essentializing Language and the Prospects for Ameliorative Projects. *Ethics* 131 (3), 460-488.

——(2021b). Essentializing Inferences. *Mind & Language*. 36 (4), 570-591.

——(2021c). Does Identity Politics Reinforce Oppression? *Philosophers' Imprint* 21 (4), 1-15.

Saul, J. (2017). Are Generics Especially Pernicious?. *Inquiry: An Interdisciplinary Journal of Philosophy*, 1-18.

Sterken, R.K. (2020). Linguistic intervention and transformative communicative disruptions. In A. Burgess, H. Cappelen, D. Plunkett (Eds.) *Conceptual engineering and conceptual ethics*. Oxford: Oxford University Press, 417-434.

Szabó, Z. G. (2015). Major Parts of Speech. *Erkenntnis* 80 (1), 3-29.

Táíwò, O. (2020). Identity Politics and Elite Capture. *Boston Review*. http://bostonreview.net/race/olufemi-o-taiwo-identity-politics-and-elite-capture

—— (2022). *Elite Capture: How the Powerful Took Over Identity Politics (And Everything Else)*. Pluto Press.

Tajfel, H., Billig, M.G., Bundy, R.P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology* 1(2), 149-178.

Thomasson, A. (2003a). Foundations for a Social Ontology. *Protosociology* 18-19, 269–290.

—— (2003b). Realism and Human Kinds. *Philosophy and Phenomenological Research* 67, 580–609.

Tirrell, L. (2012). Genocidal Language Games. In Ishani Maitra & Mary Kate McGowan (eds.), *Speech and Harm: Controversies Over Free Speech*. Oxford University Press, 174-221.

Vasilyeva, N. & Lombrozo, T. (2020). Structural Thinking about Social Categories: Evidence from Formal Explanations, Generics, and Generalization. *Cognition*. 204, 1-14.

Walton, G. M., & Banaji, M. R. (2004). Being What You Say: The Effect of Essentialist Linguistic Labels on Preferences. *Social Cognition* 22(2), 193–213.

Wierzbicka, A. (1986). What s in a noun? (Or: How do nouns differ in meaning from adjectives?). *Studies in Language*, 10, 353–389.

Wodak, D. & Leslie, S.-J. (2017). The Mark of the Plural: Generic Generalizations and Race. In Paul C. Taylor, Linda Martín Alcoff & Luvell Anderson (eds.), *The Routledge Companion to the Philosophy of Race*. Routledge. 277-289.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy* 25 (3), 287-318.

—— (2021). *Causation with a Human Face: Normative Theory and Descriptive Psychology*. Oxford University Press.

Young, I. M. (1990). *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press.

—— (1994). Gender as Seriality: Thinking about Women as a Social Collective. *Signs* 19 (3), 713-738.