An Essentialist Theory of the Meaning of Slurs[1]

Eleonore Neufeld, University of Southern California

"It's about the most insulting thing he could think of," gasped Ron, coming back up. "Mudblood's a really foul name for someone who is Muggle-born – you know, non-magic parents. There are some wizards – like Malfoy's family – who think they're better than everyone else because they're what people call pure-blood."

--- J.K. Rowling: *Harry Potter and the Chamber of Secrets*

Abstract

In this paper, I develop an essentialist model of the semantics of slurs. I defend the view that slurs are a species of kind terms: slur concepts encode mini-theories which represent an essence-like element that is causally connected to a set of negatively-valenced stereotypical features of a social group. The truth-conditional contribution of slur nouns can then be captured by the following schema: For a given slur S of a social group G and a person P, S is true of P iff P bears the 'essence' of G – *whatever* this essence is – which is causally responsible for stereotypical negative features associated with G and predicted of P. Since there is no essence that is causally responsible for stereotypical negative features of a social group, slurs have null-extension, and consequently, sentences containing them are either meaningless or false. After giving a detailed outline of my theory, I show that it receives strong linguistic evidence. In particular, it can account for a wide range of linguistic cases that are regarded as challenging, central data for any theory of slurs.

Introduction

On New Year's Eve, 2016, the Cologne Police Department proudly reported, via Twitter, that it was currently screening hundreds of "Nafris" at the main train station in Cologne.[2] The label "Nafri", used by the police to refer to people from North Africa, had its (public) linguistic debut in this tweet, which was immediately followed with national moral outrage.[3] Later, when

---

[1] This is a heavily abridged version of the original version of the paper. In the full version, I relate the essentialist theory to convergent evidence from cognitive and developmental psychology. If you are interested in reading the full version of the paper, please feel free to contact me at eneufeld@usc.edu.

[2] The original German tweet reads: "#PolizeiNRW #Silvester2016 #SicherInKöln: Am HBF werden derzeit mehrere Hundert Nafris überprüft. Infos folgen." Cf. http://www.spiegel.de/panorama/justiz/silvester-kontrollen-in-koeln-was-bitteschoen-ist-ein-nafri-a-1128172.html, downloaded 01.01.2017.

[3] The German satirist Jan Böhmermann (known for his legal conflict with Turkish president Recep Erdogan), for example, asked on Twitter: "Actually, what is the difference between Nafri and Nigger?" (German orig.: "Was ist eigentlich der Unterschied zwischen Nafri und Neger?" (cf. http://www.spiegel.de/panorama/justiz/silvester-kontrollen-in-koeln-was-bitteschoen-ist-ein-nafri-a-1128172.html, downloaded 01.01.2017).

justifying the department's choice of words, the police chief claimed: "We cannot deny the accumulation of criminal acts by persons from North African areas, and we needed to find a police internal term for that."[4] So what were people so upset about? The police department introduced a term that, according to their own assessment, functions to convey a causal link between membership in the social category of North Africans and criminal behavior. In other words, they introduced a term that *negatively essentialized* its targets: it doesn't only attribute criminal behavior to the group, it also says that members of the group have this trait *in virtue of some North-African 'nature'*. It is as if "Nafri" says: *"there is something about* North-Africans that makes them criminal". This, as I will here argue, is the key semantic characteristic of slurs. As I see it, slurs are kind terms encoding an 'essence' of a social group, which is taken to explain a number of negative features attributed to the group. In effect, then, the police department introduced a *slur* for people from North African countries into the German language community, and people were rightly upset about it.

Theories of the meaning of slurs so far offered can be broadly divided into three camps. According to the pragmatic camp, the meaning of slur terms is identical to the meaning of their neutral counterpart terms (e.g., the meaning of "faggot" is identical to the meaning of "homosexual"). This commits pragmatists to explain the derogatory power of slurs by appeal to purely extra-semantic mechanisms (cf. Anderson & Lepore, 2013; Bolinger, 2015; Nunberg, forthcoming). Next, descriptivists stipulate descriptive information into the content of slur terms, which is supposed to account for the offensive power of slurs by encoding an assertive element of contempt or the stereotype associated with the group referenced (Bach, forthcoming; Camp, 2013; Hom, 2008; Williamson, 2009). Expressivist accounts of slurs, finally, argue that slur terms

---

[4] German orig.: "Eine Häufung an Straftaten von Personen aus dem nordafrikanischen Raum lasse sich aber nicht bestreiten, und dafür müsse dann polizeiintern auch ein Begriff gefunden werden." (http://www.zeit.de/gesellschaft/zeitgeschehen/2017-01/koeln-silvesternacht-polizei-nafri-tweet-racial-profiling, downloaded 01.01.2017).

contain a speech-act like component that expresses the attitude of contempt towards the referenced group upon its usage (Hornsby, 2001; Jeshion, 2013a, 2013b; Potts, 2007; Schlenker, 2007). Despite their differences, all views have in common that they try to integrate the semantics of a given slurs' *neutral counterpart* into its content. Their meanings are identical according to pragmatists; a non-truth-conditional, expressive dimension is added to the neutral counterpart meaning according to expressivists; and some truth-conditional content is added according to descriptivists. Taken together, these views seem to exhaust all configurations in which the meaning of neutral counterparts could be used to explain the meaning of slurs.

The aim of this paper is to introduce to the philosophical discourse an essentialist theory of slurs that has, to my knowledge, not yet been given adequate consideration in the debate. Importantly, and in contrast to existing accounts, my account does *not* integrate the meaning of neutral counterparts into the semantic content of slurs. Instead, I argue that slurs are a species of *failed kind terms*; they are terms which, although introduced with the intention of designating natural kinds, fail to do so. All recognized properties of slurs are derivable from this simple semantic base; no additional extravagant linguistic entities need to be posited to account for the special features of slurring vocabulary. I begin by giving a detailed outline of my theory of the meaning of slurs in §1. In §2, I introduce key linguistic desiderata of a theory of the meaning of slurs, and show that my theory meets all of them.

§1 Slurs as failed kind terms

The main thesis of this paper is that slurs are failed kind terms; they designate an essence that is explanatorily connected to a set of negative stereotypical features of a social group. Thus, they are to be treated semantically on a par with terms such as "water", "gold" or "tiger". Scott Soames (2007) describes natural kind terms such as "water" as introduced by the following schema:

The term 'water' is to designate the unique substance of which (nearly) all members of the class of its paradigmatic samples are instances. Substances are explanatory kinds instances of which share the same basic physical constitution, which in turn explains their most salient characteristics – in the case of water samples, the fact that they boil and freeze at certain temperatures, that they are clear, potable, and necessary to life, etc. Hence, the predicate 'is water' will apply (at a world-state) to precisely those quantities that have the physical constitution which, at the actual world-state, explains the salient features of (nearly) all paradigmatic water-samples. (Soames, 2007, p. 2)

"Water", hence, designates w*hatever* underlying physical characteristic – call it '*essence*' – is shared by all "water"-members and explains and gives rise to the paradigmatic features of water. Similarly, I maintain that the slur "Nigger" designates a 'blackness essence' – whatever that is – which is causally responsible for and explains negative features stereotypically associated with being black, such as dealing with drugs or receiving welfare. "Faggot" is true of those people who share the 'gay essence' – whatever that is – which is causally responsible for and explains stereotypical negative features associated with being gay; e.g., being effeminate or carrying HIV. In general, the truth-conditional contribution of slur nouns can then be captured by the following schema: For a given slur S of a social group G and a person P, S is true of P iff P bears the 'essence' of G – *whatever* this essence is – which is causally responsible for stereotypical negative features associated with G and predicted of P.

Importantly, the claim is not that there *are* essences of the kind mentioned. Although slurs are introduced with the *intention* of designating natural kinds, in most cases, they actually fail to do so. In contrast to "water", "gold" or "tiger", there obviously will be no underlying, unified causal explanation for the set of associated features that is supposed to be explained by the essence. More concretely, there is no such thing as a 'gayness essence' which disposes male homosexuals to carry HIV or dress stylishly. There is no such thing as a 'blackness essence' which causes black people to deal drugs or receive welfare. Thus, the semantic contents of slurring words are empty.

I will now break down the structure of slur concepts[5] into three core elements that together constitute a theory-like representation encoded in those concepts. As with all kind terms, the central element of a slur is the causal component: the 'hidden unobservable' that explains and gives rise to the superficial, stereotypically observable features and actions of members of the social category in question. It is this causally deep component that we call the 'essence': it is the 'Chinese essence' of Chinese persons, the 'black essence' of black persons, or the 'gay essence' of homosexual persons. These 'essences' are to be thought of as the "underlying natures that make them the thing that they are" (Medin, 1989, p. 1476-1477), or as an object's "underlying reality or true nature that one cannot observe directly but that gives an object its identity, and is responsible for other similarities that category members share" (Gelman, 2004, p. 404). Importantly, essences can, but don't have to be, represented as being biologically grounded. In fact, it doesn't have to be known what exactly the essence is. As cognitive psychologists Medin and Ortony put it, essentialism can be thought of as "a 'placeholder' notion: one can believe that a category possesses an essence without knowing what the essence is" (Medin & Ortony, 1989).

The second component comprises stereotype features associated with the reference group, that, in contrast to ordinary natural kind terms, must be represented as *negative*. These features provide a heuristic for the identification of individuals of the essentialized group. That is, the observable surface features – which are, in the eyes of the racist, xenophobe or homophobe, dominantly negative[6] – deliver a reliable indicator for the presence of the causally powerful

---

[5] In the course of the paper, I often use the terms "term" and "concept" interchangeably. This is because I take the view for granted according to which terms inherit their linguistic meaning directly from internally individuated lexical concepts, which I understand as the smallest constituents of thought and primary bearers of meaning (Laurence & Margolis, 1999).

[6] I say "dominantly negative" because in most cases, some encoded stereotypical features are, in and by themselves, neutral. Consider, for example, skin color: obviously, this feature cannot, in and by themselves, have any negative valence. In the eyes of the racist, however, it can be a reliable indicator of the essence, and, since the essence (in the eyes of the racist) predisposes its bearer to exhibit *negative* features, also become a reliable indicator for *negative* stereotype features. In effect, an absolutely neutral surface feature such as +DARK SKIN can be, in the eyes of a racist, either a high predictor for a disposition to exhibit other negative features, or be represented as being negatively-valenced itself.

essence. And since they are caused by the essential property in question, it is assumed that members of the class have an inherent disposition to exhibit those features. Thus, it is assumed that most, but not necessarily all, individuals of an essentialized group share one or another subset of those features. But what's *decisive* for belonging to the essentialized group is the presence of the shared underlying essence or 'hidden nature'. This nature causally disposes the subject to eventually exhibit the negative surface features, whether it presently does so or not.

As the first and second component don't stand in an accidental, but in a causal-nomological relation to one another, we need a third semantic component to capture this special relation. This element is a representation of this very causal, law-like relationship. It is crucial for the informational organization of the category that is represented in our concept, since it relates the essence and the stereotypical features of the social reference group in a way that is not merely arbitrary or correlational, but grounded in causal laws.

The immensely offensive and derogatory power of slur terms and their distinctively *racist* (or xenophobic, homophobic, sexist, etc.) content directly derives from the outlined semantics. When the racist, xenophobe or homophobe applies a slur, he thereby makes the target in question – *and anyone who 'shares the same essence'* – part of the mini-theory, subjugating her to a form of causal determinism and thereby depriving her and other members of the group of human autonomy and self-determination. How the targets are disposed to act is, in the eyes of the slur user, determined and thus importantly constrained by the causally potent essence. Members of the targeted group are thus not evaluated by their individual acts or in relation to their environmental circumstances, but by (pre-)determined membership in a group. Crucially, the attributed essence is seen as disposing their bearers to act *badly*, or to exhibit *negative* features. Thus, by carrying the relevant 'group essence', the black or gay person is always predisposed to, say, be criminal or carry the HIV virus – *even if all available evidence indicates otherwise*. Taken

together, it is easy to see how the application of an essentialized slur term is derogating, demeaning and dehumanizing to the target and the entire social group she is a member of (cf. fig. 1).

Consider again the analogous behavior of other kind concepts, e.g., the natural kind concept "kangaroo". We know from cognitive and developmental psychology that young children think that kangaroos that grow up with goats will nevertheless be good at hopping. We act as if kangaroos are *just made* to hop (Gelman, 2004; Gelman & Wellman, 1991). So just as a kangaroo cannot lose its 'kangaroohood' if it is raised in a goat family, and is dispositionally 'made' to hop also if it doesn't do so (cf. for an empirical overview Gelman, 2003, 2004), so are the members of the social groups in question not evaluated by their individual circumstances or self-determined acts and decisions. This is precisely what is responsible for the dehumanizing power of slurs, as the attribution of 'essences' that pre-determine the target's dispositions, character traits, attitudes
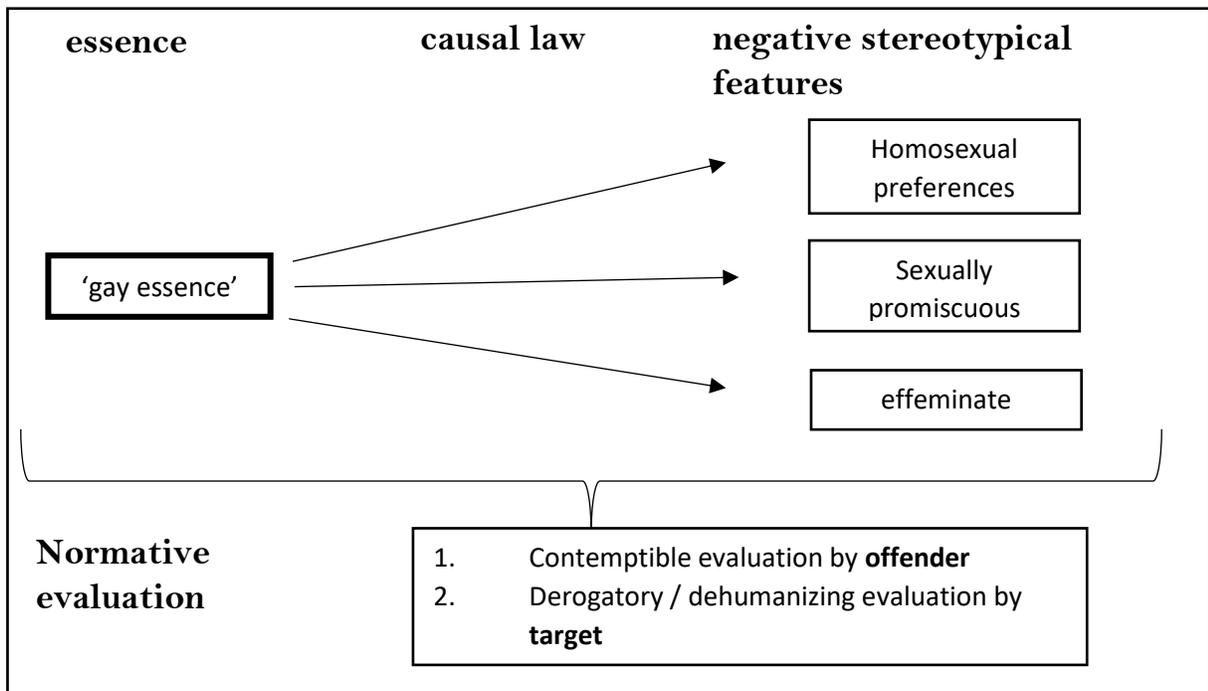

Figure 1. Model of lexical entry of a slur (exemplified by slur "faggot").

and behaviors creates a picture of the target according to which she lacks the full spectrum of human autonomy and self-determination that we associate with personhood.

Two notes before we proceed. First, let me repeat that although slurs are a species of kind terms, they possess a feature that *distinguishes* them from classic natural kind terms as "water" or "gold". In the case of "water", the essence in question is explanatorily connected to the salient features of paradigmatic instances of water; "that they boil and freeze at certain temperatures, that they are clear, potable, and necessary to life" (Soames, 2007). The valence of the salient features is completely irrelevant. In the case of slurs, in contrast, the essence must be connected to *negative* stereotypes of the social group in question. The concepts must represent that, e.g., individuals in the extension of "faggot" are emotionally sensitive and 'unmanly', dress stylishly, are sexually promiscuous and carry HIV. "Cunt" applies to individuals that are disposed to be dumb, submissive, nasty, and exclusively useful for sexual purposes. And so on: slur representations encode "Lardass" as lazy, "Kike" as greedy, or "White Trash" as having bad taste. This difference between the semantic mechanism of classic natural kind terms, on the one hand, and slur terms, on the other, accounts for the fact that slurs constitute a separate species of kind terms. Only those concepts that encode an essence which is connected to *negatively* perceived stereotypes can be called "slurs".

Second, I should emphasize that on my view, slur terms are *not* synonymous with their neutral counterparts. In fact, my account of slurs doesn't appeal to the meaning of their neutral counterparts at all. "Arab", "Jew", or "Hispanic" are governed by conventions that crucially differ from "Kanake"[7], "Kike", or "Spic". Much research confirms that race and other social concepts are highly essentialized (Cosmides, Tooby, & Kurzban, 2003; Gelman, 2003; Gil-White, 2001; Hirschfeld, 1996; Ho, Roberts, & Gelman, 2015; Rothbart & Taylor, 1992), an issue we will later

---

[7] A German slur targeted, among others, at people with Arab descent.

cover in more detail. But although these categories can be represented as socially essentialized categories that 'share a common nature' and facilitate (especially negative) generalizations (cf. Gelman, 2003; Haslam & Levy, 2006, 2006; Haslam, Rothschild, & Ernst, 2000; Hirschfeld, 1996; Leslie, 2015; Pauker, Ambady, & Apfelbaum, 2010; Prentice & Miller, 2007), the meaning of the nouns that refer to those categories is still much more innocuous, both with regard to its causal determinism and the stereotypes they encode.

First, even if we say that slurs' neutral counterparts are sometimes essentialized, the essence referred to by, say, "Chink" *is not identical to* the essence of "Chinese". Also under the assumption that ethnic labels sometimes function as kind terms, the kinds they designate differ from the kinds their closest slur-relatives designate. I merely chose 'Chinese essence' as a label for the essence placeholder that unifies, in the eyes of the racist, the alleged referents of "Chink"; I could as well have called it 'Chinkness essence'. Second, even persons whose representation of races or ethnicities is very essentialized do not have to conceptualize these racial essences as causally linked to *negative* properties in order to be competent with the race or ethnicity terms. They can believe in a 'hidden nature' of races, while not believing that the features caused by this nature are mostly negative. Thirdly, the convention governing racial (or ethnic, gender, etc.) terms generally permits higher degrees of causal innocence than the linguistic conventions governing slur terms. This means that it is *not necessary* in order to be competent with the terms "Chinese", or "Arab" to encode that instances of this kind share a causal essence that pre-disposes them to behave in negative ways. It is possible to refer to people of Arab descent in a neutral manner that *does not* essentialize them at all. In fact, many contexts require even the racist or homophobe to be aware of a non-essentialist convention that is endorsed in the case of racial, ethnic or sexual vocabulary.[8] In contrast, to fully master a slur like "faggot", 'successful

---

[8] Consider, for instance, the conventions governing legal contexts. Here, occurrences of social group terms such as "homosexual" have a purely descriptive intension whose referents can be determined by a fixed set of criteria. In this

application' *requires* one to tacitly understand the causal story between the nature of a 'gay essence' and negative stereotypes that I here outlined. In contrast to their neutral counterparts, the convention governing slurs *does not leave open* the possibility of a causally-neutral application.

§2: Slurs in Natural Language

While many theories of slurring vocabulary are already in the offering, each has been confronted with challenging linguistic data that they are not able to account for without *ad hoc* assumptions. However, as these data are acknowledged as explanatorily central for a successful theory of slurs, any adequate account should have the resources to explain them. In what follows, I will present the relevant linguistic phenomena and demonstrate that the essentialist theory can handle them in a direct, non-stipulative and uniform way.

(1) G-referencing uses of slurs

In so-called 'G-referencing'[9] uses of slurs, the slurred target belongs to the social group that is predominantly associated with the slur in question. These cases are commonly considered the most basic cases of slur usage. (1) − (3) illustrate G-referencing uses (assuming that Brian, in (2), is gay):

> (1) Let's watch the movie with those niggers in it.

> (2) Brian is a faggot.

---

case, it would be something akin to "everyone that has same-sex preferences or engages in same-sex behavior". As a result, a racist or homophobe would have to comprehend the neutral-descriptive meaning attached to the neutral counterparts in order to be competent with the terms.

[9] Robin Jeshion introduces the distinction between G-referencing, G-extending, and G-retracting uses of slurs in her extremely insightful analysis of the linguistic behavior of slurs in (2013a).

(3) The University of Southern California is full of chinks.

It is important to show that my theory does not only account for the tricky linguistic cases, but also gets the basic data right. We want to know why the application of slurs to those groups is 'licensed', and why the uses in question are derogatory. According to my theory, in each case, the targets are attributed, on the basis of some observable surface features, a Chinese, gay or black 'essence' which causally determines a set of negative features. The attribution is 'licensed' because members of the neutral counterpart group, in the eyes of the racist or homophobe, just are the paradigmatic instantiators of features that indicate the presence of the relevant essence. In almost all cases, already instantiating surface feature as +HAVING DARK SKIN, +HAVING SAME-SEX PREFERENCES, or +LOOKING CHINESE will have sufficient inductive power as to license the inference to the relevant essence for the slur user. This accounts for the meaning profile we attribute to $(1)-(3)$. The slurs apply to the targets that belong to the social group we call "neutral counterpart", because, in the eyes of the racist, the mentioned surface features license application. The uses are derogatory because making members of the social group in question – people with dark skin color in (1), homosexually-oriented people in (2), and people of Chinese ethnicity in (3) – subject to this causal-deterministic essentialization conveys to them that they don't deserve the full respect we grant persons *qua* persons.

(2) G-extending uses of slurs

Imagine the following sentence as uttered by a high school student to describe his classmate John, who doesn't like sports and has interests in art:

(4) John is not gay, but he is still a faggot.

The first thing to notice here is that intuitively, it seems to be perfectly possible to utter (4) – it does not express any contradiction.[10] However, if "gay" and "faggot" *were* truth-conditionally equivalent, as both the expressivist and pragmatist theory of slurs maintain (Anderson & Lepore, 2013; Jeshion, 2013a; Nunberg, forthcoming), (4) *should* express a semantic contradiction that can only be 'rescued' pragmatically. It would not be possible to be in the extension of "gay" without being in the extension of "faggot", and vice versa. The same holds true for descriptivist theories, as the neutral counterpart still plays a core role in the truth-conditional contribution of slurs in them (Bach, forthcoming; Hom, 2008). Take, for example, Hom's schematic characterization of the meaning of racial slurs: *ought to be subject to **these** discriminatory practices because of having **these** negative properties, all because of being NPC* (where "NPC" stands for non-pejorative correlate and 'these' points to externally realized discriminatory practices and stereotypes). However, in (4), the slurred target does *not* belong to the gay NPC. Thus, also according to descriptivists would (4) express a straightforward contradiction.

The second thing to notice is that not only is (4) usually not perceived as a contradiction, we also have immediate, clear intuitions about the information it conveys. (4) seems to convey that John is not in the extension of people with a homosexual orientation, but – because he shares salient stereotypical features associated with gay persons – is in the extension of "faggot". This poses a descriptive adequacy problem for competing theories of slurs.

Not so for the essentialist theory, for it is plain how it explains this linguistic profile. Since in my theory, slurs are not synonymous with their neutral counterparts, a contradiction is *prima facie* not predicted. This would only be so if the application of "faggot" would *entail* the application

---

[10] Note that G-extending uses of "faggot" are extremely common. As sociologist C. J. Pascoe notices in her study on masculinity and sexuality in high school, "[a] boy could get called a fag for exhibiting any sort of behavior defined as unmasculine (although not necessarily behaviors aligned with femininity): being stupid or incompetent, dancing, caring too much about clothing, being too emotional, or expressing interest (sexual or platonic) in other guys." (Pascoe, 2011, p. 57).

of "having homosexual preferences". But attribution of the slur term does not imply attribution of the neutral counterpart term. Recall fig. 1, where +HOMOSEXUAL PREFERENCES is only a surface feature of the underlying 'gay essence' cause. So the deep and hidden 'gay essence' and the superficially instantiated feature of homosexual preferences have crucially different causal roles. While it certainly has important stereotypical weight, it is possible to cancel the feature of homosexual preferences, as long as the non-changing 'gay essence', in the eyes of the slur user, 'stays present'. This is precisely what happens in (4). The homophobe's concept of "faggot" encodes a mini-theory, according to which the unobservable causal property of a 'gay essence' causes and explains observable, negative features. These stereotypical features, in turn, are the observational heuristics the homophobe uses to 'spot' the 'gay essence'. Since John presumably exhibits enough of those features, the speaker uttering (4) 'efficiently' expresses that John, although not in the extension of gay people, shares the 'essence' that causes him to exhibit other traits perceived as negative and is thus a faggot.

Robin Jeshion (2013a) dubs uses of slurs in which the slur is not applied to the counterpart group that is dominantly associated with the slur in question 'G-extending' uses of slurs. As we have seen, none of the competing models of slurs can account for G-extending uses of slurs on semantic grounds. Some theorists have tried to account for these data by stipulating that uses of slurs as in (4) are non-literal (e.g., Anderson & Lepore, 2013; Jeshion, 2013a). However, none of these theorists has offered an argument that *shows* that these uses are non-literal. In fact, it is no surprise that they advocate the non-literalness of these data, as they would falsify their theories if taken as literal uses. However, note that G-extending uses of slurs as the one in (4) are commonplace in the everyday language of slur users. Their meaning is available immediately and effortlessly, so we have *prima facie* motivation for taking these highly conventional uses at face value. The essentialist theory uniquely provides us with an independently interesting and

plausible theory that has the resources to take the data at face value and capture them without having to rely on moves that are in danger of being *ad hoc*.

(3) G-contracting uses of slurs

In so-called 'G-contracting' uses of slurs, the domain of possible targets is *contracted*: it is made explicit that the range of a slur is not the *entire* neutral counterpart group that is predominantly associated with a slur (see Jeshion, 2013a). One such example is Chris Rock's well-known line:

(6) I love black people, but I hate niggers.

Another example is

(7) My new colleague is lesbian, but she's definitely not a dyke.

As with G-extending examples, (6) and (7) are perfectly meaningful, fairly common examples of slur usage. And while G-contracting uses do not pose a problem for descriptivism, both expressivism and pragmatism have, for the reasons we encountered earlier, insufficient resources to account for the meaning ascribed to (6) and (7) on semantic grounds. Since slurs and their neutral counterparts are truth-conditionally equivalent, it is not possible to apply the neutral counterpart term to someone while denying that the target belongs to the set denoted by the slurring noun. According to the essentialist theory, the meaning of slurs and their counterparts is *not* equivalent. Thus, Chris Rock can deny that someone is a "nigger" – thus has an 'essence' that causes, e.g., criminal behavior – while attributing the surface property of dark skin to him. This is the case if the object of the discourse, in the eyes of the slur user, does not exhibit *sufficient* surface features that would license the inductive inference to the 'blackness essence'. And this fits

the intuition for (6): while the persons under discussion is black, we take the sentence to mean that they will lack many features associated with a 'black essence'.

(4) Non-derogatory, non-appropriated uses of slurs

Another species of slur that is often considered problematic in the literature are non-derogatory examples of slur uses, sometimes termed 'NDNA' uses (where 'NDNA' stands for "non-derogatory, non-appropriated"; see Hom, 2008). One example of an NDNA use is (from Hom, 2008):

(8) Institutions that treat Chinese as Chinks are morally depraved.

Accounts that treat slurs as truth-conditionally equivalent to their neutral counterpart terms struggle with the explanation of (8), since many will intuitively assign truth to (8), while rejecting (9):

(9) Institutions that treat Chinese as Chinese are morally depraved.

In addition to accounting for the disparity of truth-conditions, expressivists also struggle to account for the fact that (8) seems non-derogatory. After all, according to expressivists, each assertion of a proposition containing a slur is an expression of contempt.

By now, it should be obvious how my account explains the fact that (9) is felicitous while non-derogatory, thus meeting the adequacy condition put forward by Hom. "Chinese" and "Chinks" are not synonymous according to my account; only the latter term is true of those individuals that share a 'Chinese essence' which causes them to exhibit negative stereotypical features such as +UNHYGIENIC. The speaker of (9) thus expresses that institutions that treat the group of Chinese in this causally deterministic manner are morally depraved – which is evidently true and thus accords with our truth-intuitions.

(5) Null-extension

We have seen that an (empirical) consequence of my framework is that *most slurs have null-extension.* Scientific discovery has revealed that nothing is such that it has a determined causal 'nature' of a social group that explains and dispositionally causes the possession of stereotypically bad features. There is, of course, no such thing as a 'blackness essence', be it DNA or something else, that causes black people to receive social welfare, be criminal or sell drugs. There is no such thing as a 'Latino essence' that causes Latin Americans to rape women or work in the service industry. And so on. Slur terms are not true of anything, and consequently, sentences containing them – just as "The present king of France is bald" – are either meaningless or false. Thus, analogously to terms like "Phlogiston", slurs are examples of kind terms that have simply been *un*successfully introduced. A core condition for successful introduction of a kind term is that it is *correct* that supposed similarities of a kind have a "singly unifying explanation" (Soames, 2010, p. 89), which, in the case of slurs, is simply not given. The intuition that sentences such as

(10) There are dykes.

(11) Jews are kikes.

(12) All women are cunts.

strike us as obviously false is therefore easily captured by the essentialist theory.[11]

The null-extension consequence of my view gives us the resources to deal with a species of NDNA uses of slurs that can be classified as 'metalinguistic denial' (discussed in Jeshion, 2013a):

---

[11] Note that pragmatist and expressivist accounts of slurs make the undesirable prediction that (10) − (12) are true, since each slur is truth-conditionally equivalent to its neutral counterpart. Also descriptivist accounts that attempt to account for null-extensionality and falsehood intuitions have been confronted with challenges. For an extended discussion of various problems of Hom and May's descriptivist view with regard to null-extensionality, see Jeshion (forthcoming).

(13) There are no Chinks at my university, there are only Chinese people.

Take this sentence to be uttered by a non-racist who intends to express that the term "Chink" does not apply to Chinese people at the university, while "Chinese" does. The question is how a non-racist person could a) negate the "Chink" predicate while applying the "Chinese" predicate, and, again, b) arguably do so without expressing an offense. My framework directly predicts that (13) has these properties. The non-racist and properly informed person rejects the scientific connection between 'Chineseness' and the negative stereotypical properties encoded in "Chink". She rightly thinks that *nothing* is in the extension of "Chink": It is true of nobody that they have a 'Chineseness' nature that causally disposes them to manifest negative stereotypical features associated with being Chinese. By asserting the first conjunct of (13), she just rejects what she correctly believes to be false, as in "There are no unicorns". Since the speaker of (13) expresses that Chinese people don't fall under the extension of the term "Chink" (since no one does), we can also coin (13) as an instance of metalinguistic denial.

(6) Derogatory Projection

Importantly, although slurs have null-extension, uses of slurs still carry an *existential presupposition.* Slur users presuppose that *there are* individuals that fall under the extension of the slurs they use.[12] This accounts for a peculiar, well-known fact about slurs' projection behavior: their offensiveness persists in various compositional contexts, such as negations, conditionals, modals or questions (see Anderson & Lepore, 2013; Bolinger, 2015; Camp, 2013; Hom, 2010; Jeshion, 2013b).

---

[12] The existential presupposition is not triggered in instances of meta-linguistic denial, as in (13), or in cases of negative existentials in contexts of the type: "He isn't a Chink, no one is".

(14) He's not a Kike, he's Muslim.

(15) How many Chinks are at the University of Southern California?

(16) If he's a wop, I won't date him.

(17) She might be a dyke.

Take, as an example, (14). Although the speaker does *not* attribute a 'Jewish essence' to the object of the discourse, (14) clearly stays an instance of offensive slur usage that derogates the entire social group of Jewish people.

Usually, when we introduce entities into a discourse by talking about them, we signal to our interlocutors that we take their existence for granted:

(18) Did you feed the cat?

(19) This isn't silver, it's stainless steel.

(20) If that's lemonade, I want it.

(21) All dogs are scared of fireworks.

Needless to say, uttering (18) − (21) presupposes that you believe that there are cats, silver, lemonade or dogs. Analogously, utterance of any sentence in (14) − (17) is only felicitous if the speaker presupposes that such things as "Chinks", "Kikes", "dykes" or "wops" exist. But presupposing, like the speaker does in (14), that there *is* something like a 'Jewish essence', carried predominantly by Jewish people, that causally predisposes them to exhibit negative features, of course dehumanizes and derogates the entire group of Jewish people.[13] This explains why the

---

[13] Note that we can successfully apply the well-known 'wait a minute' test to (14) − (17), revealing the existential presuppositions triggered by the examples. This test is standardly employed to test the presuppositions triggered by a sentence (von Fintel, 2009). Consider,

(22) Stephen stopped smoking.

derogatoriness of slurs persists even if the speaker does not assertively predicate a causally potent essence to the discourse object. And since in reality, nothing is in the extension of these slurs, the informed speaker is licensed to respond with a denial of the presupposed content. Correspondingly, our earlier example (13) would be an appropriate answer to the question asked in (15), as it expresses refusal to accept the existential presupposition "There are Chinks" that the speaker introduces by the utterance of (13).

(7) Derogatory Variation

Let us now turn to the last phenomenon. It is widely-thought that some pejoratives are more powerful in their disparaging and derogating force than others, a phenomenon that is standardly listed as a central explanatory desideratum for theories of slurs (Anderson & Lepore, 2013; Hom, 2008, 2010). Compare, for example, the difference in offensiveness between the terms "Nigger" and "Limey" – the former is substantially stronger in its derogatory force than the latter (cf. Hom, 2008).[14] The same holds for "Chink" vs. "Kraut", "Kike" vs. "Honky", "Wog" vs. "Yank", and so on. Furthermore, the derogatory content of a slur can vary as a function of time: the derogatory force of "Kraut" or "Commie", for instance, was substantially stronger during the time of World War II and the Cold War, respectively, than it is now.

Many theories contend that the difference in encoded negative attitudes or negative descriptive information is what accounts for the fact that slurs differ in their derogatory strength. *Prima facie*, this explanation seems very plausible. For example, we often find that powerful slurs

---

The 'wait a minute' test reveals that (17) presupposes that Stephen smoked. If I am not willing to accommodate the common ground appropriately – because I think that Stephen never smoked – I can felicitously respond: "Wait a minute – Stephen never smoked!". This test can be successfully applied to (14) – (17). E.g., "Wait a minute – there are no Kikes!" is an appropriate answer to (12), conveying the refusal to accommodate the common ground as required by the existential presupposition.

[14] See Mullen and Leader (2005) and Rice et al. (2010) for an empirical quantification of these differences.

are also associated with very negative stereotypes. Most would agree that the negative stereotypes associated with whites ("Honkie") are less negatively valenced than the stereotypes associated with people of Chinese ethnicity ("Chink"), which in turn are less negative than the ones associated with black persons ("Nigger"). And knowing that others think ill of us or harbor negative attitudes against us *hurts*. This is true even when the agents that harbor these attitudes are not in any relevant way significant to us. Imagine you notice how a bunch of teenagers in the subway are snickering, making it obvious that you are the source of their amusement. These teenagers are complete strangers, and you will never see them again. You know that whether these teenagers think well of you or not has no impact whatsoever on anything you take to matter in your life. Still: their snicker hurts, and quite usually so.[15] Naturally, then, stronger negative attitudes will hurt more, and weaker negative attitudes will hurt less.

However, this explanation can't be the entire story. It misses out on a general, systematic pattern of how the derogatory force of different slurs varies. Why is it that in general, slurs that target someone on the basis of their race, ethnicity, gender or sexuality are particularly toxic? It is hard to imagine a slur targeting fans of, say, an opposing football team to possibly be more derogatory than slurs such as "faggot", "Kike", "Chink", or "Nigger" – even if the properties associated with these fans were highly negative. By the same token, slurs like "Kike", "faggot", or "Nigger" seem to be more diminishing than "Lardass", "Libtard", or "riffraff", although the corresponding groups are, from the perspective of the users, associated with highly negative stereotypes.[16] Theories that solely rely on differences in negative attitudes or descriptive information can't account for this fine-grained pattern of the data.

---

[15] This example leans on a similar example from Stephen Bero's 2017 USC dissertation.
[16] The high negative valence of the stereotypes associated with these social groups is supported by abundant research; see, e.g., Cvencek, Nasir, O'Connor, Wischnia, and Meltzoff (2015); Copping, Kurtz-Costes, Rowley, and Wood (2013); Cvencek, Meltzoff, and Greenwald (2011); Bessenoff, G. R., & Sherman, J. W. (2000); Devine (1989); Gaertner and McLaughlin (1983); Dovidio, Evans, and Tyler (1986); Woods, Kurtz-Costes, and Rowley (2005); Wang, Brownell, and Wadden (2004).

The essentialist theory captures this subtle pattern. The slurs we find particularly toxic – the ones targeting someone on the basis of their race, sexuality, gender, or ethnicity – are the ones which, in addition to encoding profoundly negative stereotypes, are *highly essentialized*. In each case, the slur expresses that it is in someone's *very nature* to have features that are *bad*. And while knowing that you think badly of me hurts, knowing that you think badly of me because of something in my inescapable nature is *deep*. What I do and who I am is not seen as a matter of my individual choices and agency, but as a deep matter of my nature. This is what it means to dehumanize.

Let us go in more detail through my semantic model and the way it explains the data. I maintain that the derogatory force of a slur is a direct offspring of its semantics, where the essence *and* the set of negative features are the determining factors of a slurs' meaning. The derogatory strength of a slur therefore is a function of these elements. The more negative the represented stereotype of a group is, the more demeaning the corresponding slur should come out.[17] The more a category is essentialized, the stronger the diminishing effects of the slur should be. When the two of them come together, the derogatory force of slurs is explosive.[18] Thus, holding the level of essentialism (largely) fixed, slurs for groups with strong negative stereotypes are more derogatory ("Lardass" compared to "Wheat thin", "Chink" compared to "Kraut", or "Commie" during the Cold War as opposed to now).[19] Holding the degree of negative stereotyping fixed,

---

[17] Of course, the represented stereotype must also *conventionally govern* the slur in question. However, we can say that our representation of the stereotype of the social group most associated with the slur is a rough measure of the stereotype convention that governs the slur.

[18] Interestingly, high level of essentialism towards a social group has often been found to *predict* negative stereotyping and prejudice (see Bastian & Haslam, 2006; Haslam & Ernst, 2002; Haslam & Levy, 2006; Howell et al., 2011; Leslie, 2015; Levy & Dweck, 1999; Pauker et al., 2010; Prentice & Miller, 2007; Williams & Eberhardt, 2008). In addition, Sarah-Jane Leslie (2015) argues that there's a tendency to attribute features to an essence of a basic-level category when these features are negative.

[19] These comparisons are not completely clean – it proves hard to keep the essentialist or stereotype dimensions fixed when making comparisons. It is very plausible, for example, that the social category of communists was more essentialized during the Cold War than it is now. Similarly, although 'race' receives generally the highest essentialism ratings, essentialism for the category 'black' is still more marked than essentialism for 'white' (see Haslam et al., 2000).

slurs for groups which are strongly essentialized will be more derogatory (e.g., "White Trash" vs. "riffraff", or "Kraut" vs. "Leftie"). When a category is *both* strongly essentialized and the associated stereotypes are highly negative, the diminishing force of a slur culminates (as, e.g., in "Cunt", "faggot", or "Nigger").

While it is evident that the negative stereotypes we associate with a group can be more or less pronounced, it has also long been established that there are differences in the degrees to which we essentialize social groups (see Prentice & Miller, 2007, p. 202). In an important study conducted by Nick Haslam and his colleagues (2000), they developed a set of questions that assessed different dimensions along which we essentialize groups. Specifically, they tested whether participants essentialized social groups along the dimensions of naturalness, stability, discreteness of category boundaries, immutability of category membership, and necessity of category features.[20] Within the 40 social categories that were rated,[21] the categories of *gender, ethnicity, and race* as well as *Jews* and *homosexuals* received particularly high ratings, and the categories associated with interests, politics, appearance, and social class received the lowest ratings.[22] And if we bring to our minds the slurs that are the derogatorily deepest, we will directly see that they fall under one of those social categories.

---

[20] They took these to be the dimensions "that are commonly invoked in psychological, philosophical and social scientific writings" (Haslam, Rothschild, & Ernst, 2000). In the study, they asked participants to rate, on a scale from 1 to 9, to which extent one of the listed dimensions applies to a category. The dimension of discreteness, for example, was described to the participants as follows: "Some categories have sharper boundaries than others. For some, membership is clear-cut, definite, and of an "either/ or" variety; people either belong to the category or they do not. For others, membership is more "fuzzy"; people belong to the category in varying degrees". To give another example, the dimension of naturalness was described to the participants as "Some categories are more natural than others, whereas others are more artificial".

[21] The aim of the study was to cover as many categories as possible; some of the assessed categories were, e.g., diseases (AIDS patients, cancer patients), intelligence groups (smart people, people of average intelligence), races (black, white) religions (Catholics, Jews), political groups (liberals, Republicans).

[22] It is important to clarify that the categories *Jews* and *homosexuals* received extremely high ratings for dimensions that Haslam et al. describe as indexing a group's *entitativity*. Entitativity is a subtype of our general essentialist bias, and can be described as the belief that members of a group are very similar to one another, such that membership in a group is very informative about the nature of its members – in other words, the group is seen as *entitative*. Essentialism, as entitativity, has been found to predict prejudice and negative stereotyping towards groups such as homosexuals.

In sum, theories that appeal to differences between descriptive or expressive attitudes towards different social groups only can account for derogatory variation if we carve up the data in a very coarse-grained, one-dimensional way. To account for the subtle pattern we find in the data, we need another parameter. The essentialist theory delivers this level by adding another dimension to the derogatory potential of a slur: the derogatory force is a function not only of the negative stereotypes it encodes, but its stereotypes *and* the degree to which it essentializes. The essentialist theory, then, uniquely captures the systematic patterns we find in the data and explains why slurs that are particularly deep in their offensiveness tend to fall under specific categories; categories that are strongly essentialized.

Conclusion

In this paper, I defended a novel account of the semantics of slurs. Slurs, I argued, are akin to kind terms in that they denote an essence of a social category which nomologically connects to a set of negative stereotypical features. I demonstrated that essentialism about slurs explains the recognized linguistic properties of slurs, thus demonstrating its explanatory reach. My goal here was to make a cumulative case for my theory. Although someone might disagree with the assessment of some of the data, and a thorough review of how competing theories handle the data is a task that I can't do justice in this paper alone, it is important to note that my view does not stand or fall on the basis of how we assess a single piece of evidence. The main attraction of a theory stems from its resources to predict and account for a *wide* range of data. I believe the last section has illustrated that the essentialist theory does precisely that (see table 1 for an overview).

Data Positions	G-referencing	G-expanding	G-contracting	NDNA-uses	Null-extension	Derogatory Projection	Derogatory Variation
Pragmatism	✓	✗	✗	✗	✗	✗	✗
Expressivism	✓	✗	✗	✗	✗	✓	✗
Descriptivism	✓	✗	✓	✓	Not clear	Not clear	✗
Essentialism	✓	✓	✓	✓	✓	✓	✓

Table 1. Comparison of data-reach of main theories of slurs

References

Anderson, L., & Lepore, E. (2013). Slurring Words. *Noûs*, *47*(1), 25–48. https://doi.org/10.1111/j.1468-0068.2010.00820.x

Bach, K. (forthcoming). Loaded words: On the semantics and pragmatics of slurs: to appear in. In David Sosa (Ed.), *Bad Words*. Oxford: Oxford University Press.

Bessenoff, G. R., & Sherman, J. W. (2000). Automatic and controlled components of prejudice toward fat people: Evaluation versus stereotype activation. *Social Cognition.* (18(4)), 329–353.

Bolinger, R. J. (2015). The Pragmatics of Slurs. *Noûs*, n/a-n/a. https://doi.org/10.1111/nous.12090

Camp, E. (2013). Slurring Perspectives. *Analytic Philosophy*, *54*(3), 330–349. https://doi.org/10.1111/phib.12022

Copping, K. E., Kurtz-Costes, B., Rowley, S. J., & Wood, D. (2013). Age and Race Differences in Racial Stereotype Awareness and Endorsement. *Journal of applied social psychology*, *43*(5), 971–980. https://doi.org/10.1111/jasp.12061

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in cognitive sciences*, *7*(4), 173–179. https://doi.org/10.1016/S1364-6613(03)00057-3

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development*, *82*(3), 766–779. https://doi.org/10.1111/j.1467-8624.2010.01529.x

Cvencek, D., Nasir, N. S., O'Connor, K., Wischnia, S., & Meltzoff, A. N. (2015). The Development of Math-Race Stereotypes: "They Say Chinese People Are the Best at Math". *Journal of Research on Adolescence*, *25*(4), 630–637. https://doi.org/10.1111/jora.12151

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology, 22*(1), 22–37. https://doi.org/10.1016/0022-1031(86)90039-9

Fintel, K. von. (2009). Would you believe it? The king of France is back! Presuppositions and truth-value intuitions. In A. Bezuidenhout & M. Reimer (Eds.), *Descriptions and beyond.* Oxford [u.a.]: Oxford Univ. Press.

Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly*, 23–30.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*: Oxford University Press, USA.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in cognitive sciences, 8*(9), 404–409.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition, 38*(3), 213–244. https://doi.org/10.1016/0010-0277(91)90007-Q

Gil-White, F. (2001). Are ethnic groups biological "species" to the human brain? *Current anthropology, 42*(4), 515–553.

Haslam, N., & Ernst, D. (2002). Essentialist Beliefs About Mental Disorders. *Journal of Social and Clinical Psychology, 21*(6), 628–644. https://doi.org/10.1521/jscp.21.6.628.22793

Haslam, N., & Levy, S. R. (2006). Essentialist beliefs about homosexuality: structure and implications for prejudice. *Personality & social psychology bulletin, 32*(4), 471–485. https://doi.org/10.1177/0146167205276516

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39*(1), 113–127. https://doi.org/10.1348/014466600164363

Haslam, N., Rothschild, L., & Ernst, D. (2004). Essentialism and Entitativity: Structures of Beliefs about the Ontology of Social Categories.

Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and the child's construction of human kinds*: MIT Press.

Hom, C. (2008). The semantics of racial epithets. *The Journal of Philosophy, 105*(8), 416–440.

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

Hornsby, J. (2001). Meaning and uselessness: how to think about derogatory words.

Howell, A. J., Weikum, B. A., & Dyck, H. L. (2011). Psychological essentialism and its association with stigmatization. *Personality and Individual Differences, 50*(1), 95–100. https://doi.org/10.1016/j.paid.2010.09.006

Jeshion, R. (2013a). Expressivism and the offensiveness of slurs. *Philosophical Perspectives, 27*(1), 231–259.

Jeshion, R. (2013b). Slurs and stereotypes. *Analytic Philosophy, 54*(3), 314–329.

Jeshion, R. (forthcoming). Embracing Corruption: A Response to Hom and May. In David Sosa (Ed.), *Bad Words.* Oxford: Oxford University Press.

Leslie, S.-J. (2013). Essence and natural kinds: When science meets preschooler intuition. *Oxford studies in epistemology, 4*(108), 9.

Leslie, S.-J. (2015). The original sin of cognition: Fear, prejudice, and generalization. *The Journal of Philosophy, 3.*

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. *Similarity and analogical reasoning, 179*, 195.

Mullen, B., & Leader, T. (2005). Linguistic Factors: Antilocutions, Ethnonyms, Ethnophaulisms, and Other Varieties of Hate Speech. In J. F. Dovidio, P. S. Glick, & L. A. Rudman (Eds.), *On the nature of*

prejudice. Fifty years after Allport (pp. 192–207). Malden, MA: Blackwell Pub. https://doi.org/10.1002/9780470773963.ch12

Nunberg, G. (forthcoming). The Social Life of Slurs. In D. Fogal, D. Harris, & M. Moss (Eds.), *New Work on Speech Acts*. Oxford University Press.

Pascoe, C. J. (2011). *Dude, you're a fag*. Berkeley.

Pauker, K., Ambady, N., & Apfelbaum, E. P. (2010). Race salience and essentialist thinking in racial stereotype development. *Child Development, 81*(6), 1799–1813. https://doi.org/10.1111/j.1467-8624.2010.01511.x

Potts, C. (2007). The expressive dimension. *Theoretical linguistics, 33*(2), 165–198.

Prentice, D. A., & Miller, D. T. (2007). Psychological Essentialism of Human Categories. *Current Directions in Psychological Science, 16*(4), 202–206. https://doi.org/10.1111/j.1467-8721.2007.00504.x

Rice, D. R., Abrams, D., Badea, C., Bohner, G., Carnaghi, A., Dementi, L. I.,. . . Trzebinski, J. (2010). What Did You Just Call Me? European and American Ratings of the Valence of Ethnophaulisms. *Journal of Language and Social Psychology, 29*(1), 117–131. https://doi.org/10.1177/0261927X09351696

Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds?

Schlenker, P. (2007). Expressive presuppositions. *Theoretical linguistics, 33*(2), 237–245.

Soames, S. (2007). What are natural kinds? *Philosophical Topics, 35*(1/2), 329–342.

Soames, S. (2010). *Philosophy of language*: Princeton University Press.

Wang, S. S., Brownell, K. D., & Wadden, T. A. (2004). The influence of the stigma of obesity on overweight individuals. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity, 28*(10), 1333–1337. https://doi.org/10.1038/sj.ijo.0802730

Williamson, T. (2009). Reference, Inference, and the Semantics of Pejoratives. In J. Almog & P. Leonardi (Eds.), *The philosophy of David Kaplan* (pp. 137–158). Oxford, New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367881.003.0009

Woods, T. A., Kurtz-Costes, B., & Rowley, S. J. (2005). The Development of Stereotypes About the Rich and Poor: Age, Race, and Family Income Differences in Beliefs. *Journal of Youth and Adolescence, 34*(5), 437–445. https://doi.org/10.1007/s10964-005-7261-0