An Essentialist Theory of the Meaning of Slurs[1]

Eleonore Neufeld, University of Southern California

Forthcoming in *Philosophers' Imprint*
(penultimate draft; please cite final version)

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, many sentences containing them are either meaningless or false. After giving a detailed outline of my theory, I show that it receives strong linguistic support. In particular, it can account for a wide range of linguistic cases that are regarded as challenging, central data for any theory of slurs. Finally, I show that my theory also receives convergent support from cognitive psychology and psycholinguistics.

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

---

[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.

"Nafri", used by the police to refer to North Africans, had its (public) linguistic debut in this tweet, which was immediately followed with national moral outrage.[3] Later, when justifying the department's choice of words, the police chief claimed that "It is undeniable that there is an 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 its 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.

The main aim of this paper is to introduce to the philosophical debate an essentialist theory of slurs that has, to my knowledge, not yet been given adequate consideration. The main thesis of this paper is that slurs[5] are a species of *failed kind terms*; they are terms which, although

---

[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).

[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).

[5] One question that has come up in the literature on slurs is how to delineate the class of slurs, and, relatedly, how to distinguish slurs from other pejoratives (see, e.g., Anderson & Lepore, 2013; Bach, 2018; Croom, 2011; Jeshion, 2013a; Popa-Wyatt, 2016). My answer to this question has direct consequences for the scope of my theory. Although I agree with other theorists that there is a clear distinction between paradigmatic slurs and paradigmatic non-slurring pejoratives (e.g., so-called 'individual pejoratives' like "jerk" or "asshole"), I disagree with them by holding that these two classes lie on a *continuum* rather than being clearly definable. Correspondingly, the boundary of the scope of my

introduced with the intention of designating kinds, fail to do so. All recognized properties of slurs are derivable from this simple semantic base; no additional linguistic entities need to be posited to account for the special features of slurring vocabulary. Although the primary goal of this paper is to motivate an essentialist semantics of slurs, rather than to defeat theoretical alternatives, it is worth mentioning that I take my essentialist model to have a central virtue that makes it stand out from competing theories. Namely, that it can account for the acknowledged desiderata of an adequate semantics of slurs while receiving strong support from empirical work in cognitive psychology.

In what follows, I assume a theory of natural kind terms according to which they encode an essence of a kind, k, that is explanatorily connected to a set of stereotypical features associated with k. This way of carving out the semantics of natural kind terms differs slightly from the classical Kripkean framework of natural kind terms. Since I take descriptive information about stereotypical features to be part of the lexical entry of kind terms, I am committed to a conception of lexical representations as informationally rich. However, instead of using this paper to debate issues about the informational structure of lexical representations, I will here simply assume this framework and show what some of its fruits are.[6]

I begin by giving a detailed outline of my theory. Next, I introduce key linguistic desiderata of a theory of the meaning of slurs, and show that my theory meets all of them. Finally, I present evidence from cognitive psychology and psycholinguistics for my essentialist account.

---

theory is fuzzy. In the section on derogatory variation, I will explain how the mechanics of my model *explain* that there are middle cases that are not clearly classifiable into either category (e.g., "fatso", "leftie").

[6] In fact, my results will be the same under a framework of natural kind terms according to which they rigidly designate an essence and do not encode any stereotypical features. Since these are used to descriptively fix the reference upon introduction of the kind term, the reference of slur terms will still be empty, but the alleged referents will still be presupposed to possess an 'essence' with negative causal-dispositional potential.

§1 Slurs as failed kind terms

1.1   The View

The main thesis of this paper is that slurs are akin to natural kind terms. Under the framework

of natural kind terms I am assuming, natural kind terms were introduced to designate an essence

that is explanatorily connected to a set of stereotypical features of a kind. Slur terms are

distinctive because they designate[7] an essence that is explanatorily connected to a set of *negative*

stereotypical features of a social group. Thus, slurs are a species of kind terms and 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 *whatever* 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 N-word is used to designate 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 gay persons; e.g., that they are effeminate or carry

---

[7] Throughout the paper, I do *not* use "designate", "refer", and cognate expressions as success terms. As has already
become clear, the view defended in this paper is that slurs have null-extension, and thus do not designate or refer to
anything.

HIV.[8] In general, slur concepts encode mini-theories which represent an essence-like element that is causally connected to a set of negatively-valenced stereotypical features associated with 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.

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 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 into three core elements that, according to the view that I am advocating here, together constitute a theory-like representation encoded in those concepts.[10] The central element of a slur is the causal component: the intrinsic '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'. These 'essences' are to be thought of as the intrinsic,

---

[8] After the term has been introduced into a language community, it is possible that some stereotypes associated with a slur change. For example, at the time of introduction of "faggot", the 'gayness essence' wasn't taken to explain the feature of carrying HIV, since the slur predates the discovery of the virus. In these cases, we simply discover *more* features to be caused by the essence (just as with other natural kind terms), and update the concept accordingly. Insofar as the updated concept is sufficiently similar, concept identity is preserved (see section 1.3). Note, however, that the focus of this paper is on lexical, not diachronic, semantics. In-depth discussion of diachronic phenomena, such as acquisition, lexical transition from non-slurring to slurring meaning, meaning identity over time, or appropriation have to be addressed in a separate paper.

[10] 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.

"underlying natures that make them the thing that they are" (Medin, 1989, 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, the essence can, but does not have to be, represented as being biologically grounded. In fact, it does not 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).[11]

The second component comprises stereotype features of 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[12] – deliver a reliable indicator for the presence of the causally powerful 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

---

[11] Thus, the notion of "essentialism" that I operate with is the one that is used in the literature on psychological essentialism (Gelman, 2003; Haslam et al., 2004; Medin & Ortony, 1989), which, in the philosophical literature, is sometimes referred to as "quintessentialism" (Leslie, 2013).

[12] The negativity-aspect of the theory raises an important question: What does it take for a feature to be negative? Generally, I take a quasi-subjectivist stance on this matter: a feature is negative when it is represented as negative by a sufficient number of subjects. Under this conception, negativity is highly context-sensitive. Even if a feature is generally seen as positive or neutral, it can become negative in certain contexts. For example, while +HAVING HIGH SAT SCORES or +DRESSING STYLISHLY are, in and by themselves, positive qualities, they are evaluated as threatening and negative when combined with certain social outgroups, as in the first example, or certain genders, as in the latter. Similarly, many encoded features, such as skin tone or facial configuration, will be objectively absolutely valence-free but can either be encoded as proxies for other negative features and thus themselves become represented as negative, or be, irrationally, encoded as negative in the first place.

A reviewer also pointed out to me that the negativity-aspect of my proposed semantics might create a problem in accounting for sentences such as "I'm hopeless at fashion. I wish I knew some fag who could just tell me how to dress to attract the ladies", which should come out infelicitous under my account, since it is used to ascribe and explain properties the speaker takes to be positive. However, I do not think this is correct. We often use negative properties of others to our favor, as in: "I'm in love with this woman, but she has a husband. I wish I knew some criminal, bad person who could help me get rid of my problem." This is perfectly coherent, despite the fact that +CRIMINAL or +BAD are negative properties and the speaker represents them to be negative – the person just uses these negative properties to his favor. For a persuasive response to a similar objection by Camp (2013), see Jeshion (2018).

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 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 derogatory, toxic power of slur terms and their distinctively *racist* (or xenophobic, homophobic, sexist, etc.) content directly derives from the outlined semantics. When
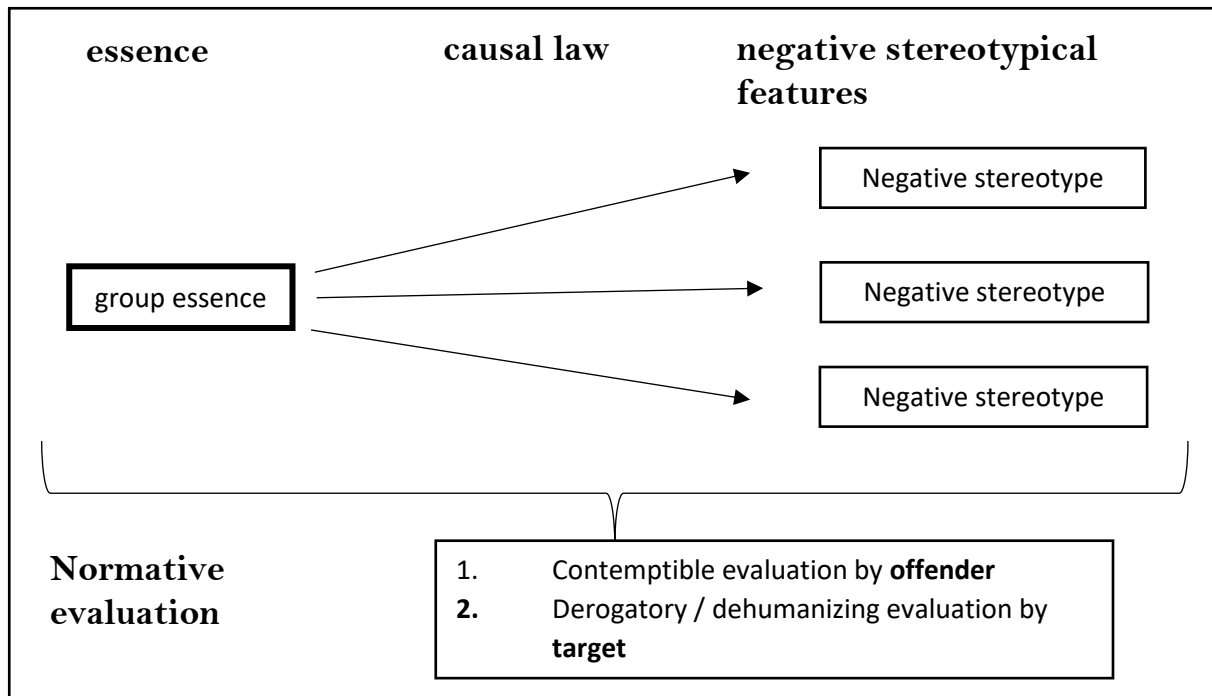

Figure 1. Model of lexical entry of a slur.

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 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.[13] 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, for example, 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

---

[13] See Basu (2019) for a recent argument that epistemically representing others in a way that treats them as scientific objects – i.e., essentializes them – constitutes a case of wronging.

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.[14]

1.2   Some Helpful Contrasts

One of the most important things to emphasize is 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. "Gay", "Jew", or "Hispanic" are governed by conventions that crucially differ from "faggot", "kike", or "spic". Much research confirms that race 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 cover in more detail. But although "Muslim", "black", or "Chinese" can be represented as socially essentialized categories that 'share a common nature' and facilitate (especially negative) generalizations (cf. Dovidio, Glick, & Rudman, 2005; Gelman, 2003; Haslam & Levy, 2006; Haslam, Rothschild, & Ernst, 2000; Hirschfeld, 1996; Leslie, 2015b; 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.

---

[14] Remember that although slurs are a species of kind terms, they possess a feature that *distinguishes* them from classic natural kind terms as "water", "gold" or "tiger". 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. According to the convention governing, e.g., "faggot", they are emotionally sensitive and 'unmanly', dress stylishly, are sexually promiscuous and carry HIV. The convention governing "cunt" is that they are exclusively useful for sexual purposes. And so on: the convention governing "Lardass" is that they are lazy, the one of "Kike" that they are greedy, and "White Trash" have, according to linguistic convention, 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 and, correspondingly, a separate linguistic category. Only those concepts that encode an essence which is connected to *negatively* perceived stereotypes can be called "slurs".

First, even if we say that slurs' neutral counterparts are sometimes essentialized, I contend that 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 kind they designate differs from the kind their closest slur-relative designates. 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 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", "black", or "homosexual" 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 with an Arab background 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.[16]

To explain the mechanics of my view, it is helpful to draw contrasts to some other, superficially similar views. My account bears similarities to the hybrid family-resemblance account by Adam Croom (Croom, 2011, 2014a, 2015) and the perspectival account by Elisabeth

---

[16] 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 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. In contrast, to fully master a slur word, 'successful application' *requires* one to tacitly understand the causal story between the nature of 'black 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.

Camp (Camp, 2013, 2018). According to Croom, slurs contain both an expressive and a descriptive component, the latter of which consists of a list of weighed prototypical features. Since the prototypical features encoded by slurs and their neutral counterparts differ, it follows that they are not truth-conditionally equivalent. Although my account, like Croom's, treats slurs as informationally rich, there are a number of important differences between them. Under my account, slurs do not merely encode feature lists; rather, they encode information in a way that is *causally organized.* In particular, slurs relate prototypical features nomologically to causally powerful essences.[17] As we will see in the next sections, the causal elements explanatorily differentiate my account from Croom's, for they lead to a number of distinct linguistic and psychological predictions about phenomena such as derogatory variation, essentialism about social groups, or nominalization. Most importantly, the causal elements play a crucial role in ensuring that slurs will come out as empty, as there is and will be no 'deep essence' that explains features associated with a group, *even if the stereotypes, due to effects of occurrent structural disadvantages, will be true of certain members of oppressed groups.*[18]

Similarly, Camp argues that "slurs make two distinct, coordinated contributions to a sentence's conventional communicative role: a truth-conditional predication of group membership, and endorsement of a derogating perspective on that group" (Camp, 2018, p. 30). In virtue of the second speech act, a speaker signals their allegiance to a *perspective*[19], according to

---

[17] For defenses of causally-structured models of concepts and criticisms of purely feature-based models of conceptual representation, see Danks, 2014; Gelman, 2003; Murphy, 2004; Rehder, 2017; Sloman, 2005; Sloman, Love, & Ahn, 1998.

[18] Notice another difference: To explain the relationship between neutral counterparts and slurs, Croom posits a 'conceptual anchor', individuated by the prototype property with the largest feature weight. These features have a "grounding role" insofar as they "serve as salient anchors for the semantic or conceptual content of slurs" (Croom, 2015, p. 35). Although properties corresponding to the neutral counterpart will often have high inductive weight and thereby explain the relationship between slurs and paradigmatic targets, these features, on my account, do *not* 'ground' or 'anchor' the meaning of slurs, just like +WET does not ground the meaning of "water".

[19] Camp characterizes a perspective as "an intuitive tool for structuring thoughts: a disposition to notice, explain, and respond to some part of the world in certain ways. Perspectives in general may, but need not, include any particular propositional or attitudinal commitments; and they are partly, but only partly, under voluntary control" (Camp, 2018, p. 30).

which the target's group membership is *explanatory* of many of her other properties, and predicts the display of negative *stereotypical* properties. This second speech-act is similar in spirit to the semantics I propose here.

Despite these similarities, there are important differences that differentiate our accounts. Perhaps most importantly, I only posit one, purely predicative, speech-act to explain the semantics of slurs.[20] As the pure truth-conditional attribution of group membership doesn't play any role on my account, Camp's first, predicational speech-act comes out as explanatorily redundant on my account. As a result of this difference, the accounts diverge with regard to some key linguistic predictions, which we will assess in the next section. As with Croom, one of the crucial predictive differences is that predications of slurs will always come out false under my account, whereas they often come out as true within the predicative speech-act on Camp's account.

But also the second, perspectival speech-act does not do what slurs do under my construal. According to Camp, "slurs are akin to other expressions [like "tu"/"vous" or slang expressions for parents, food, or genitals], part of whose conventional function is not merely to refer or predicate, but to signal the speaker's social, psychological, and/or emotional relation to that semantic value" (Camp, 2013, p. 335); thus, slurs contribute *"a (broadly) expressive, perspectival element to the conversation"* (my emphasis; Camp, 2018, p. 48). The latter quote is instructive: the second speech-act is broadly expressive, because it is *about* the speaker's perspective on a referent. In contrast, although they can reveal something about my perspective – just as calling a chair "sofa" can reveal something about my perspective on the chair – slurs are not *about* perspectives on my account. In my view, slurs' meaning is predicative in the full-fledged, traditional sense. What slurs *say* of you is that you have some group essence that disposes you to display bad

---

[20] Another difference is that Camp explicitly rejects that slurs, generally, conventionally encode stereotypes. However, because she suspects that some slurs do encode stereotypes, I will treat this difference as not too important (see Camp, 2013).

features; thus, a slur-predication will be either true or false of you. It is precisely because of the purely predicative function that slurs come out as empty.

1.3   *Objections*

Before we move on, let me respond to a salient objection against the proposed model, namely, that it is too informationally demanding for competent language speakers. Two worries seem particularly concerning. First, two people can plausibly employ a slur in communication without talking past each other, although each of them associates different stereotypes with it. Second, someone can be competent with a slur without knowing the stereotype associated with it. In the slurs literature, this objection has been spelled out by Robin Jeshion and Elisabeth Camp (Camp, 2013; Jeshion, 2013b). However, it is important to bear in mind that these objections are in fact versions of familiar, more general worries about rich views of lexical meaning that have come up *outside* of the slurs debate, such as inferential role or prototype theories of concepts (Fodor, 1998; Fodor & Lepore, 1992; Rey, 1983).

The main goal of this paper is to assume – not defend – an account of lexical meaning as richly structured and argue that it helps us explain certain patterns that are unique to slurs. The background semantic framework I assume for this end is the basic paradigm behind a family of views, all of which are still live options in the current theoretical discourse on the nature of meaning and conceptual structure.[21] So although I will briefly respond to the objections that have

---

[21] Issues regarding the richness of conceptual structure and its intersection with linguistic competence constitute an object of ongoing, vivid discussion in philosophy and the cognitive sciences (Block, 1987; Del Pinal, 2016, 2018; Fodor, 1998, 2005; Fodor & Pylyshyn, 2014; Gädenfors, 2000; Harman, 1993; Jönsson, 2017; Kamp & Partee, 1995; Knobe, Prasada, & Newman, 2013; Lakoff, 1987; Leslie, 2015; Marconi, 1997; Prinz, 2012; Putnam, 1975; Soames, 2015; Taylor, 2009).

come up against treating slurs along these lines, the appropriate locus for a response is not this paper, but a paper discussing the *general* viability of this approach to meaning and concepts.

A number of philosophers and cognitive scientists have issued powerful replies against the first objection (see, e.g., Chomsky, 2000; Harman, 1993; Marconi, 1997; Smith, Medin, & Rips, 1984). Their strategy emphasizes that, holding the level of competence fixed, *similarity of conceptual content* is all we need to explain the stability of meaning between different speakers, and, for that matter, communicative success. More concretely, if the mental concepts that two speakers associate with a word are sufficiently similar, we would expect that information exchange, in most cases, proceeds smoothly.[22] Appealing to high similarity instead of strict identity also explains the undeniable fact that we sometimes miscommunicate or are in disagreement about the extension of a given term in borderline cases. If the meaning of every word was strictly identical between each competent speaker, these phenomena would become a mystery. Thus, modeling meaning stability in terms of content similarity explains communicative success and exceeds the descriptive accuracy of a strict identity view. This point can directly be applied to slurs. Due to similarity of content, communication will proceed 'smoothly' in most cases. Only in rare borderline cases, communication between two subjects might be unsuccessful due to encoded feature differences.

Is it possible to be competent with the meaning of a slur and not have knowledge of any associated stereotype? According to Jeshion, someone can coherently and competently utter

---

[22] This point is quite important and often neglected. It belongs to the *operationalization* of a stereotype that *it is highly stable*: something is a stereotype *only if it is highly stable* in a community. If there was no cross-subject stability, we would not call a given property a *stereotype*. Relatedly, if stereotypes allowed for high variability, we would not get reliable and replicable effects in experimental paradigms and hence wouldn't be able to find any stereotype effects in controlled settings -- but we do. Also detrimental phenomena such as the stereotype threat would not be very worrisome if the associations triggered were as variable across subjects as is suggested by the objection.

(1) "I disdain those queers; anyone who would do that is sick. But I do not endorse those [stereotypes] as the right ways of thinking about queers. I have no idea who does it, what they are like, and I don't care. I just think those queers should be locked up."[23] (Jeshion, 2013b, p. 322)

Granting that the intuition Jeshion capitalizes on is generalizable,[24] an essentialist perspective still affords a couple of moves explaining it away. The speaker in (1) has to do quite some work and introduce a variety of qualifications to get their intended, minimal meaning of "queer" across. As it is, a richly-structured theory suits quite well to let them do precisely that: rich theories of concepts are famously associated with more fine-grained and contextually dynamic conceptions of meaning, and thus allow for flexibility of meaning and extension. Although the speaker rejects many stereotypes, one surface property that the speaker regards as highly negative is left, which is made salient through "anyone who would *do* that". According to my theory, the context of utterance modifies the meaning in such a way as to make clear that the speaker means to pick out someone who bears the 'queerness essence' which causally explains the one negative property of same-sex behavior.[25] And although we can make sense of this contextually modified case in a way fully compatible with the essentialist theory, the *standing*

---

[23] Note that Jeshion does not control for a case in which a speaker disavows commitment to the feature of engaging in same-sex behavior, which seems to be as felicitous as her example:
"I disdain those queers, any guy who dresses effeminate is sick. I do not endorse the same-sex way as the right way of thinking about them. I have no idea who does it, what they are like, and I don't care. I just think those queers should be locked up."

[24] As has been pointed out before (Camp, 2013), it is unclear how uniform the intuitions about (1) are, and thus whether our theory should accommodate this data point. I, for one, have extreme difficulties making sense of (1), especially if schematically replaced by other slurs. Here's one reason that might explain my difficulty. It seems to be a true generalization that slurs emerge in communities that *interact* with the people they are slurring. This is why slurs are often fairly meaningless to people who are from cultures or communities that are not in touch with the slurred group in question. But if this is true, it will also be true that, due to interaction with slurred groups, these communities and competent users within them *will* have stereotypes of these members. It is therefore fairly difficult to imagine that anyone who is competent with a slur could utter something like (1). Notice that Jeshion herself goes on to explain that "Much racism and bigotry is rooted simply on finding others "different" – often because of physical characteristics" (p. 322). +DIFFERENT or +PHYSICAL CHARACTERISTIC C, however, *are* stereotypical properties of the same status as the stereotypes Jeshion addresses (e.g., +SEXUALLY PROMISCUOUS).

[25] And possibly a number of other negative properties that the speaker leaves open.

meaning of "queer" will still be one that fully corresponds to the semantics I propose – i.e., one in which more than only the minimal stereotype is communicated.[26]

Drawing on Putnam's division of linguistic labor (Putnam, 1975), we can furthermore appeal to *partial* linguistic competence and deference to experts to accommodate the general worry under which Jeshion's is subsumable. Can we say that Putnam is linguistically competent with the word "elm", even if his associated prototypes of "elm" and "beech" are identical? In some sense, yes: he knows that "elm" is a natural kind term and designates some 'elm essence' that explains common properties of elms. He will not be at total loss in linguistic discourse about elms, and he will have an idea about the reference of the term when elms are nearby. He himself will also be able to apply the term correctly in many circumstances. However, Putnam also knows that there are *degrees of competence*, and that there are "elm" experts in his linguistic community whose referential and inferential competence with the term exceed his. Thus, Putnam is disposed to take the "elm" expert as a linguistic authority when the circumstances require, and revise his concept in accordance with the expert's more fine-grained one. Although Putnam can be said to be competent with the word "elm", he certainly does not have the expert's degree of linguistic competence.

This point can be applied to (1). Suppose that a speaker is ignorant about most stereotypes speakers associate with the slur. All they know is that it is used towards people with the surface property +HOMOSEXUAL. Can we say that the person is competent with the slur? Again, in some sense, yes. Since they know that the term is a slur, they know that it denotes an essence that must explain the feature +HOMOSEXUAL and some other features. Thus, in many circumstances, they

---

[26] Analogously, we can say something like: "I love guns. But I'm not endorsing shooting as the right way of thinking about guns. I just think anything made of plastic and sold in Toys'R'Us is great for my kids." Again, due to the awkward context, the speaker, with quite some work on behalf of her and her listener, communicates that her extension is restricted to guns that don't shoot people, but are made for children. However, this doesn't show that usually, the meaning of "gun" doesn't involve the property of shooting.

will not be at loss in linguistic discourse deploying the term, draw inferences from it, and be able to make out intended referents. But again, since competence comes in degrees, we wouldn't say that the speaker is *fully competent*: there will be situations in which they will be confused about the referent of the word – for example, in cases in which the intended referent of a more competent user does not engage in same-sex behavior, but exhibits other stereotypes. Since also the speaker knows that there are "experts" of the term in their linguistic community, they will be disposed to update their lexical entry of the slur if they take their interlocutors to be linguistic authorities.

§2: Slurs in Natural Language

Having presented the view, I will now demonstrate its explanatory reach. Slurs exhibit unique, peculiar linguistic patterns that have proven difficult to capture. However, as these data are acknowledged as explanatorily central for a successful theory of slurs, any adequate account must have the resources to explain them. The data include: (1) G-extending, (2) G-contracting, and (3) G-referencing uses of slurs (Croom, 2015; Jeshion, 2013a), (4) non-derogatory, non-appropriated uses of slurs (Hom, 2008; Jeshion, 2013a), (5) intuitions about null-extension and falsehood (Hom, 2008; Hom & May, 2013; Richard, 2008; Sennet & Copp, 2015), (6) projection behavior (Anderson & Lepore, 2013; Camp, 2013, 2018; Cepollaro & Stojanovic, 2016; Jeshion, 2013b; Potts, 2007) and (7) derogatory variation of slurs (Bolinger, 2017; Hom, 2008; Jeshion, 2013b; Nunberg, 2018). In what follows, I will go through these linguistic phenomena, and demonstrate that the essentialist theory can handle them in a direct, non-stipulative way.

(1) 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:

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

Similarly, we can imagine someone dissenting with (3a), yet assenting with (3b):

(3) a. "John is gay."

b. "John is a faggot."

The first thing to notice here is that intuitively, it seems to be perfectly possible to utter (2) (or to disagree with (3a) yet accept (3b), for that matter) – intuitively, it does not express any contradiction.[28] However, if "gay" and "faggot" *were* truth-conditionally equivalent, as is defended in many prominent accounts on slurs, (2) *should* express a semantic contradiction that can only be 'rescued' pragmatically. For example, Anderson and Lepore's minimalist analysis treats slurs' semantic content to be exhausted by the truth-conditional semantic content of the neutral counterpart (Anderson & Lepore, 2013). Also Jeshion's expressivist account treats slurs' truth-conditional contribution to be equivalent to the truth-conditional contribution of the neutral counterpart; an expressive element of contempt is added to account for the derogatory properties of slurs (Jeshion, 2013a). And as we saw earlier, according to Camp, one of the speech-acts slur users engage in when they use a slur is a pure predication of membership in the neutral counterpart group (Camp, 2013, 2018). Thus, in all these cases, it would not be possible to be in the extension of "gay" without being in the extension of "faggot", and vice versa.[29] The second

---

[28] 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).

[29] The same point applies to the analyses in Bach (2018); Cepollaro & Stojanovic (2016); Hom (2008); Hom and May (2013, 2015); Hornsby (2001); Nunberg (2018); Potts (2004); Schlenker (2007); Sennet and Copp (2015); Whiting (2013); and Williamson (2009), because they either include the meaning of the neutral counterpart into the truth-

thing to notice is not only that (2) is usually not perceived as a contradiction, we also have immediate, clear intuitions about the information it conveys. Namely, that John is not in the extension of people with a homosexual orientation, but – probably because he shares salient stereotypical features associated with gay persons, such as talking about 'female' topics, being physically 'weak', or dressing stylishly – is in the extension of "faggot".

The essentialist theory can handle the examples in (2) and (3) quite smoothly. Consider the homophobe uttering (2). Since in my theory, slurs are not synonymous with their neutral counterparts, a contradiction is not predicted. This would only be so if the application of the slur would *entail* the application of "having homosexual preferences". But attribution of the slur term does *not* imply attribution of the neutral counterpart term. Recall fig. 1: +NEGATIVE STEREOTYPE X is only a surface feature of the underlying 'group essence' cause. Thus, 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 (2). 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 (2) 'efficiently' expresses that John, although not in the extension of gay people, shares some gayness 'essence' that causes him to exhibit negative traits correlated with gayness and thus falls under the extension of the slur.

---

conditional meaning of a slur, or treat slurs and neutral counterparts as truth-conditionally equivalent (conditional on the existence on a neutral counterpart term—see Nunberg, 2018).

Robin Jeshion (2013a) dubs cases as the ones in (2) and (3) – in which the slur is applied to a target that doesn't belong to the group paradigmatically associated with the slur – 'G-extending' uses of slurs.[30] Some theorists have tried to explain away G-extended uses by stipulating that they are non-literal (e.g., Anderson & Lepore, 2013; Jeshion, 2013a); therefore, a theory of slurs need not account for these uses. However, none of these theorists has offered an *argument that shows* that these uses are non-literal. The usual move is to point to other cases of metaphorical language-use and assume that these cases and G-extending uses of slurs are parallel. However, *whether* these uses are parallel is precisely what is at stake – I yet have to see a defense of this claim that establishes, and not simply assumes, that they are.[31] In fact, it is no surprise that these theories advocate the non-literalness of these data, as they would falsify their theories if treated as literal uses. And although it is not my claim that this strategy is principally unavailable to debunk the data in (2) and (3), it is important to note that the essentialist theory 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 treat G-extending uses as non-literal.[32]

Note also that G-extending uses of slurs as the one in (2) are commonplace in the everyday language of slur users.[33] Their meaning is available immediately and effortlessly, so there is *prima facie* strong motivation for taking these highly conventional uses to be literal.[35] Additional

---

[30] 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).

[31] The cleanest test I can think of in which two quite uncontroversially truth-conditionally equivalent open class expressions in different word forms are employed in a way similar to (2) is "Guillermo is not a bachelor, but he is an unmarried man." This example, however, does *not* pattern with (2) or (3). It is incredibly hard, if not impossible, to make sense of the statement.

[32] Also Cepollaro (2017) notices that it is never argued for, but always assumed, that G-extending uses and other metaphorical language uses are parallel.

[33] See fn. 28.

[35] One might think that this data point would also support an analysis of slurs in terms of conventional implicature (e.g., Copp, 2009; Whiting, 2013; Williamson, 2009); however, note that the conventional implicature analyses on offer will still predict that (1) and (2) result in contradictions.

evidence comes from constructions with modifiers and qualifiers such "true", "real", and "deep down":

> (4) "Although Leyla isn't a socialist, she's still a true/real commie."

> (5) "Although Jack isn't black, he's still a true/real nigger."

> (6) a. "I know that he's not gay, but deep down, he's a faggot."

> b. "I know you haven't lied, but deep down, you're a liar."

Intuitively, taking the slur usages in (4) to (6a) to be non-literal seems implausible. This point is strengthened if we look at the use of "liar" in (6b). In (6b), which parallels (6a), what the speaker is *literally* accusing me of is to be, deep down, a liar. To say that "liar" is used in a non-literal sense seems unjustified: to be a liar is precisely what the speaker is accusing me of and thus accountable for. Similarly, it is difficult to make sense of the claim that the speaker in (6a) labels someone as a "faggot" in any non-literal way – after all, the speaker claims that this is what the targeted person, deep down, *is*. The intended meanings of (4) – (6) are furthermore directly available, even though the objects of discourse don't belong to the neutral counterpart groups in questions *and* the slurs have been combined with literality modifiers, such as "true" and "real". These data, while not absolutely decisive, seriously undermine the claim that G-extending are non-literal uses of slurs.

(2) 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):

> (7) "I don't have anything against feminists – in fact, I'm a feminist myself. What I hate are these feminazis."

(8) "Although my best friend is gay, you can be sure that he's not a faggot."

(9) "Thank God! My new neighbors are Lesbians, but they are not dykes."

As with G-extending examples, (7) to (9) are perfectly meaningful, fairly common examples of slur usage.[36] As before, a number of accounts predict that this sentence yields a semantic contradiction, since they subscribe to the view that slurs and their neutral counterpart are truth-conditionally equivalent.[37] 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.[38] According to the essentialist theory, the meaning of slurs and their counterparts is *not* equivalent. Thus, a slur user can deny that someone has an 'essence' that causes negative properties while attributing the bare property of homosexual preferences 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 'gayness essence'.[39] And this fits the intuition for (8): while the person under discussion has homosexual preferences, we take the sentence to mean that he will lack many features associated with a 'gay nature'.

---

[36] For a selection of other examples as evidence against co-referentialism about slurs, see Croom (2015).

[37] As in the last section, advocates of the hypothesis that slurs and their neutral counterparts are truth-conditionally equivalent can resort to the claim that the slur uses in (7) to (9) are non-literal. Again, I don't claim that there are no strategies for these theories to accommodate this data point – after all, pragmatic moves are always available to explain *any* phenomenon away. However, leaving aside the reasons I gave in the last section against a pragmatic and in favor of a semantic analysis, I'd like to re-emphasize that the essentialist theory accommodates this phenomenon smoothly without *relying* on additional pragmatic explanations and is therefore arguably explanatorily more powerful.

[38] In fact, accounts under which semantic content is exhausted by truth-conditional content (e.g., Anderson & Lepore, 2013; Nunberg, 2018) predict that the meaning of G-contracting sentences is identical to their G-extending counterpart sentences, since ⟦slur⟧ = ⟦counterpart⟧. This prediction hasn't been made explicit in the literature to my knowledge, but strikes me as fairly worrisome.

[39] Rehder (2017) gives a comprehensive overview of how essentialist concepts can be modelled probabilistically in terms of causal graphical models; given the weight attached to surface features, we can use Bayes' Theorem to model causal inference from surface features to essence in all applications of slurs (including G-referencing, G-extending and G-contracting uses).

(*3*) G-referencing uses of slurs

In so-called 'G-referencing' 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. $(10) - (12)$ illustrate these G-referencing uses:

(10) "Let's watch the movie with those cunts in it."

(11) "Of course, the kraut made me fail the exam."

(12) "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, German or female '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 such as +LOOKING FEMALE or +HAVING GERMAN ETHNICITY 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 $(10) - (12)$. 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: making members of the social group in question subject to this causal-deterministic

essentialization conveys to them that they don't deserve the full respect we grant persons *qua* persons.

(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 'non-weapon' (Jeshion, 2013a) or 'NDNA' uses (where 'NDNA' stands for "non-derogatory, non-appropriated"; see Hom, 2008). One example of an NDNA use is (from Hom, 2008):

(13) "Institutions that treat Chinese as chinks are morally depraved."

Accounts that treat slurs as truth-conditionally equivalent to their neutral counterpart terms predict that (13) is true exactly when (14) is:

(14) "Institutions that treat Chinese as Chinese are morally depraved."

But intuitively, many speakers would assign truth to (13) while rejecting (14). In addition, many expressivist accounts furthermore predict that (13) is derogatory, since each assertion of a proposition containing a slur is an expression of contempt.[40] However, according to some theorists – prominently, Chris Hom – (11) is an instance of a non-derogatory (even if upsetting, triggering and hence offensive) speech-act.[41]

---

[40] A reviewer pointed out to me that NDNA uses of this kind could be analyzed metalinguistically. Although I understand – in fact, as will become clear in the next sections, advocate – this point when applied to some cases of negation (as has been done, for example, in Anderson & Lepore, 2013; Bolinger, 2015; Jeshion, 2013) and particularly rigorously in Cepollaro, 2017), it is hard for me to see how a metalinguistic move can be applied to (13). Especially in light of the fact that (a) no slur is mentioned before the occurrence of the slur that a metalinguistic use could pick up on and (b) no reading in which a comment on a linguistic item is made is available to me, since (c) "as" clearly introduces an intensional context, it is unclear to me how a metalinguistic analysis for this case should work. But since the main focus in this part of the paper is to motivate the essentialist theory by showing that it covers a wide range of data, and not so much on refuting alternative theories, I will leave this issue aside for now.

[41] Like others, I have the intuition that even uses in intensional contexts like (13) will be upsetting or offensive, e.g., as a result of triggering effects. However, we might still want to agree with Hom that there is some principled

By now, it should be obvious how my account explains the fact that (13) is felicitous and at least different in derogatory status from the examples we encountered earlier. "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 (13) 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 – importantly, the ones we regard as particularly toxic – *don't have any extension* in the actual world. In these cases, 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 'latino essence' that dispositionally causes Latin Americans to rape women or work in the service industry. There is no such thing as a 'blackness essence' that dispositionally causes black people to receive social welfare or sell drugs. There is no such thing as a 'gayness essence' that dispositionally causes homosexual men to carry HIV or be emotionally sensitive. And so on. Slur terms are not true of anything, and consequently, sentences predicating slurs of individuals 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

---

difference in degree of *derogation* (as opposed to offensiveness) between non-NDNA uses and NDNA uses, and expect a theory to capture this difference. Alternatively, the slur might trigger an existential presupposition such that derogation projects out even in this intensional context, in which case the phenomenon would fall under my later discussion of derogatory projection.

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

(15) "There are dykes."

(16) "Jews are kikes."

(17) "All women are cunts."

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

One question that arises here is whether essentialism about slurs really secures null-extension. Oppressed, socially-constructed groups can be subject to unjust practices that, given the world as it unjustly is, connect group membership in non-accidental ways to negative properties that emerge from the oppression. Since the oppressive forces converge on individuals because they are taken to be a member of the group, the connection is causal. Since membership in the socially constructed kind is in part a question of whether one is taken to be or treated as a member of that kind, the causal connection is through group membership. If that is the case, the conditions specified by a slur might be satisfied and the slur successfully refers.[42] This is an important objection to the null-extension hypothesis. In response, I want to highlight certain assumptions that constrain our representations of 'essences': we see them as *internal* and *intrinsic* to the subject; a class of subjects cannot possess an essence relationally. As Haslanger (2011) puts it for the case of generics: "[generics license the inference that] the fact in question obtains by virtue of something *specifically about the subject so described*, i.e., about women, or blacks, or sagging pants. In the examples I've offered, however, this implication is unwarranted. The facts in question obtain by virtue of broad system of social relations within which the subjects are situated, and are not grounded in *intrinsic or dispositional features of the subjects themselves*." (my emphasis,

---

[42] Thanks to an anonymous reviewer for raising this point.

Haslanger, 2011, p. 446). Similarly, the causal element of slurs presupposes that the essence is intrinsic, not extrinsic, to the subject. Since this condition is not satisfied in the cases I'm discussing, slurs don't have extension.

In contrast to my account, many other accounts, such as Anderson and Lepore's minimalist account and Jeshion's expressivist accounts, are committed to the view that the sentences in (15) to (17) are true. But also accounts that lie closer to the account I propose here differ with the predictions they make about slurs' reference and, correspondingly, the truth of sentences containing slurs. Consider Croom's family resemblance view of slurs (Croom, 2011, 2015). According to him, slurs encode a set of negatively-valenced weighted features. If a person P satisfies sufficiently many features of the feature list associated with a slur S, S is *true* of P. However, it is not implausible that there will be individuals that happen to exhibit the features associated with the slur – as a matter of structural injustices, or simply because of their very individual life choices. Thus, the extension of slurs will not be the empty set under Croom's account. Specifically, a sentence such as (18) would have to be treated as strictly true

(18) "Alberto is a spic."

if Alberto happens to satisfy a number of features associated with the slur "spic", which, according to Croom, include features such as "x is a Mexican-American" or "x is a foreign worker or exchange student with a thick non-native accent" (Croom, 2014b, p. 162). The essentialist account differs from Croom's insofar as what is decisive for the successful reference of the slur is whether the target possesses some intrinsic 'Latino essence', which slur users take to unify all "spics", that disposes them to exhibit negative traits. Thus, (18) will be false even if Alberto happens to exhibit a number of features corresponding with the stereotype. Again, while these accounts might be

able to appeal to explanations that lie outside the domain of their theories to explain our falsehood and referential intuitions about slurs away,[44] the essentialist theory accounts for them directly.

The null-extension consequence of my view also gives us the resources to deal with a species of NDNA uses of slurs that can be classified as 'metalinguistic denial':

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

Take this sentence to be uttered by a non-racist who, upon hearing (12), 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 slur predicate while applying the "Chinese" predicate, and, again, b) do so without derogation.[45] My framework predicts that (19) has these properties. The non-racist and properly informed person rejects the causal connection between a 'Chineseness essence' 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 Chinese. By asserting the first conjunct of (19), she just rejects what she correctly believes to be false, as in "There are no unicorns". Since the speaker of (19) expresses that Chinese people don't fall under the extension of the term "Chink" (since no one does), we can also classify (19) as an instance of metalinguistic denial.

Before we move on, I will address an objection that Sennet and Copp (2015) raise against the null-extensionality thesis as defended in Hom and May (2015). If successful, it would also apply to my version of the thesis. Fortunately, it isn't. Their charge is that null-extensionality of slurs entails that sentences of the following kind are trivially true:

---

[44] For such a strategy see, e.g., Anderson and Lepore (2013); Camp (2018); Jeshion (2013b) or Whiting (2013).
[45] See fn. 41.

(20) "All kikes are Mormons."

But, contra null-extensionality, Sennet and Copp remark that (20) is intuitively false.

A standard position in formal semantics is that we should introduce an existential domain condition (in other words, a *lexical existential presupposition*) to our semantics for universal quantifiers,[46] since this would increase the descriptive accuracy of our theory of quantifiers with respect to sentences like $(21) - (23)$, which would all come out as trivially true without such a condition.

(21) All mermaids live in Ohio.

(22) Every unicorn admires Noam Chomsky.

(23) All phlogiston is located in the Pacific Ocean.

If you agree with the position that quantifiers come with lexical existential presuppositions, then (20) will come out as false or truth-valueless. If you don't agree with it, (20) will come out as trivially true, but so will (21) to (23) – which doesn't lead anyone to worry about whether "mermaid", "unicorn", or "phlogiston" in fact have an extension. In sum, the objection doesn't pose any problem for null-extensionality views of slurs.

(6) Derogatory Projection

Importantly, although slurs have null-extension, we have just seen that uses of slurs still carry an *existential presupposition*.[47] Slur users presuppose that *there are* individuals that fall under the

---

[46] See, e.g., Diesing (1992), McCawley (1972), Strawson (2012). See Heim and Kratzer (1998), ch. 6 for an introductory overview of the debate. The lexical entry of "all" would change to $\lambda < A, B > : A \neq \emptyset. A \subseteq B.$

[47] Note that there is no controversy about whether sentences like $(20) - (23)$ introduce *some* existential presupposition. The question is whether it is to be located lexically or pragmatically.

extension of the slurs they use.[48] This accounts for a peculiar, well-known fact about slurs' projection behavior: their derogatory effect persists in various compositional contexts, such as negations, conditionals, modals or questions (Anderson & Lepore, 2013; Bolinger, 2017; Camp, 2013, 2018; Cepollaro & Stojanovic, 2016; Hom, 2010; Jeshion, 2013b).

(24) "He's not a kike, he's Muslim."

(25) "How many chinks are at the University of Southern California?"

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

(27) "She's so bad with the wand, she might be a mudblood."

Take, as an example, (24). Although the speaker does *not* attribute a 'Jewish essence' to the object of discourse, (24) clearly stays an instance of derogatory slur usage. Why?

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

(28) Did you feed the cat?

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

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

(31) All dogs are scared of fireworks.

Needless to say, uttering (28) − (31) presupposes that you believe that there are cats, silver, lemonade or dogs. Analogously, utterance of any sentence in (24) − (27) is only felicitous if the speaker presupposes that there *are* objects in our domain the slurs apply to. But presupposing, like the speaker does in (24), that there is something like a 'Jewish essence', carried predominantly

---

[48] The existential presupposition is not triggered in instances of metalinguistic denial, as in (19), or in cases of negative existentials in contexts of the type: "He isn't a chink, no one is".

by Jewish people, that causally predisposes them to exhibit negative features, of course dehumanizes and derogates the entire group of Jews.[49] This explains why the derogatoriness of slurs persists even if the speaker does not assertively predicate a causally potent essence to a discourse object. And since in reality, nothing is in the extension of slurs, the informed speaker is licensed to respond with a denial of the presupposed content. Correspondingly, our earlier example (19) would be an appropriate answer to the question asked in (25), as it expresses refusal to accept the existential presupposition that the speaker introduces by the utterance of (19).

(7) Derogatory Variation

Let us now turn to the last explanandum on our list. 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; Bolinger, 2017; Hom, 2008, 2010). Compare, for example, the difference in offensiveness between the N-word and "Limey" – the former is substantially stronger in its derogatory effects than the latter.[50] The same goes 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.

---

[49] Note that we can successfully apply the well-known 'wait a minute' test to (24) – (27), 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,

(32) Stephen stopped smoking.

The 'wait a minute' test reveals that (24) 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!". Similarly, "Wait a minute – there are no kikes!" is an appropriate answer to (16), conveying the refusal to accommodate the common ground as required by the existential presupposition.

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

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.[51] *Prima facie*, this explanation seems very plausible. For example, we often find that powerful slurs are also associated with very negative stereotypes. Most would agree that the negative stereotypes associated with white people ("honky") 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 (N-word). And knowing that others think ill or harbor negative attitudes against us *hurts*. This is true even when the agents that harbor these attitudes are not 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.[52] 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?[53] 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" or "kike" – even if the properties associated with these fans were highly negative. By the same token, these slurs seem to be more diminishing than "lardass", "libtard", or "junkie", although the corresponding groups are, from the perspective of the users,

---

[51] That holds true especially of views that are close to the view I advocate here (Croom, 2011; Hom, 2008), but is also a move open to expressivist views.
[52] The example is based on an example used in Bero (2017).
[53] In a similar vein, Popa-Wyatt asks in her 2016 paper: "Is it coincidence that many of the most offensive slur words are associated with groups we might identify as oppressed?" (Popa-Wyatt, 2016, p. 155). My answer to that question is 'no': group essentialism is the key variable explaining both group oppression and the derogatory potential encoded in slurs (Appiah, 1996, 2018; Leslie, 2017; Livingstone Smith, 2011).

associated with highly negative stereotypes.[54] Theories that solely rely on differences in negative attitudes or descriptive information can't account for this fine-grained pattern of the data.

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 intrinsic, 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.[55] 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.[56] Thus, holding the level of essentialism (largely) fixed, slurs for groups with stronger negative stereotypes are more

---

[54] 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 & Sherman (2000); Devine (1989); Gaertner and McLaughlin (1983); Dovidio, Evans, and Tyler (1986); Woods, Kurtz-Costes, and Rowley (2005); Wang, Brownell, and Wadden (2004).

[55] 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 associated with the slur is a rough measure of the stereotype convention that governs the slur.

[56] 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). See also Sarah-Jane Leslie (2017) for a discussion of our tendency to attribute features to an essence of a basic-level category when these features are negative.

derogatory.[57] Holding the degree of negative stereotyping fixed, slurs for groups which are more essentialized will be more derogatory. When a category is *both* strongly essentialized and the associated stereotypes are highly negative, the diminishing force of a slur culminates.

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.[58] Within the 40 social categories that were rated,[59] 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.[60] 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.

---

[57] 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).

[58] They took these to be the dimensions "that are commonly invoked in psychological, philosophical and social scientific writings" (Haslam et al., 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".

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

[60] 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*.

Note also that the essentialist theory gives us a natural way to accommodate the theoretical difficulty of finding a clear demarcating criterion distinguishing slurs from non-slurs. Many theorists draw a distinction between slurs, which target individuals based on their membership in a group, and individual pejoratives, which target individuals based on some (temporary) behavior or "personal qualities". While everyone can point at paradigmatic examples of slurs, and paradigmatic cases of individual pejoratives ("jerk", "asshole", "dickhead"), and most feel the intuitive pull to theoretically distinguish between these two classes (although see Jeshion 2013 for a criticism of this distinction) there are many pejorative terms that have proven to be quite difficult to classify in one way or the other. Consider,

(26) "Hey fatso!"

Popa-Wyatt notices that (26) – like "lardass", "druggie", "junkie", "bum" or "commie" – "sits in the middle ground between [slurs] and [individual pejoratives] [...]. Like pejoratives targeted at individuals, they identify the targeted individual on the basis of specific properties that s/he has. But like slurs, they express contempt not only about the particular individual but also about other people who have similar features, and so may be identified as part of a group" (2016, p. 152). Essentialism about slurs explains why it is difficult to find a clear line distinguishing slurs from individual pejoratives. Just as essentialism ratings are on a continuum, our judgements about whether something is a slur or not will be on a continuum, rather than an all-or-nothing affair. People are less sure about whether people with higher body weight or communist attitudes are describable by having some 'group essence' that determines their behavior and unifies them, or whether they should be characterized simply as having mutable, individual properties. Hence, in

---

Essentialism, as entitativity, has been found to predict prejudice and negative stereotyping towards groups such as homosexuals (Haslam & Levy, 2006).

these cases, subjects will be reluctant, unsure or in disagreement about whether to call a term "slur" or an "individual pejorative", just as predicted by the essentialist theory.

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 of a slur 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.

§3 Slurs and the Psychology of Social Kinds

3.1 Experimental Evidence for the Essentialist Theory

Thus far, I have motivated my theory by showing that an essentialist semantics for slurs can account for their main linguistic properties. I now present converging evidence from cognitive psychology for the existence of the cognitive structures postulated by my theory of slurs. As I indicated earlier, I here assume an intimate link between linguistic meaning and mental concepts. More specifically, according to the view I take for granted in this paper, 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. Thus, granting this assumption, studying the structure of and information encoded in lexical concepts can directly inform our semantic theory. We will review evidence in favor of the two central components of

my essentialist analysis: (1) slurs are semantically organized in an essentialist way, and (2), slurs are uniquely associated with negative stereotypes.

Let us start with component (1). Through a number of well-established psychological paradigms, cognitive psychologists have documented that certain categories – natural kind categories such as animals, minerals, and chemicals – are cognitively represented in a highly essentialized way. We act as if members of certain categories have immutable, enduring, and natural essences which make them what they are (for an overview, see Gelman, 2003; 2004). Interestingly, however, as we have already seen in the last section, we now know that we also think of a large variety of *human* or *social categories* in this exact same, highly essentialized, way (Gelman, 2003; Gil-White, 2001; Haslam, 2000; Haslam & Levy, 2006; Haslam et al., 2000; Hirschfeld, 1995, 1996; Prentice & Miller, 2007).[62] More concretely, we behave as if social groups are *real kinds*: they have sharp category boundaries, are somewhat 'natural', historically stable, 'real' and not constructed, or allow for rich inductive inferences about physical and behavioral traits of their members (Demoulin, Leyens, & Yzerbyt, 2006; Haslam et al., 2000).[63] In particular, social categories such as race and ethnicity (Gil-White, 2001; Haslam et al., 2000; Hirschfeld, 1996; Ho et al., 2015; Verkuyten, 2003), gender (Gelman, 2003; Gelman, Collman, & Maccoby, 1986; Prentice & Miller, 2006, 2007), caste (Mahalingam, 2003), sexual orientation (Haslam & Levy, 2006), mental illness (Haslam, 2000; Haslam & Ernst, 2002; Howell, Weikum, & Dyck,

---

[62] There is wide-ranging evidence that we hold essentialist beliefs from early childhood on, which have been documented by psychologists – prominently, Susan A. Gelman – throughout the past 30 years. For example, pre-school children believe that a baby kangaroo raised among goats will grow up to hop and have a pouch (Gelman & Wellman, 1991; cf. Gelman, 2003). They also expect that something that has turtle insides will still be a turtle even if it doesn't look like one (Gelman & Wellman, 1991), suggesting that they don't rely on observable surface features to determine kind membership. For an overview of the evidence tracking children's essentialist belief structure, see Gelman (2003).

[63] It is important to keep in mind that these markers are *characteristic features* of our representation of essences, and do not constitute necessary conditions for something to be represented as an essence. This conception of essences also doesn't completely correspond to the philosopher's as "that intrinsic aspect of a thing which grounds all and only the intrinsic metaphysical necessities that hold of the thing" (see Leslie, 2015 for the same point).

2011) – in short, the categories that are central to human slurring practices – have been found to be cognitively represented in a highly essentialized way.[64]

For illustration, consider the case of race: this category is certainly among the most relevant for slurs, given both the prevalence and deep offensiveness of epithets that target subjects on the basis of their race. In a series of pivotal experiments, psychologist Lawrence Hirschfeld documented essentialist thinking about race in both adults and preschoolers as young as three years (Hirschfeld, 1995, 1996; see also Pauker et al., 2010). In one paradigm, he asked preschoolers whether a racial property class – hair and skin color – or a physical property class – clothing style and color – would remain unchanged as a person grows up. Even 3-year-olds judged that the properties connected with race were more constant than sartorial properties.[65] He obtained the same preference for race as the dominant factor compared to other physical features for inheritance judgements: when children were asked which properties they would inherit from their parents, they predominantly picked racial properties. In a switched-at-birth paradigm, children were asked which racial properties a child that was adopted by parents of another skin color would develop. 5-year olds outweighingly decided in favor of the birth parents' racial properties (Hirschfeld, 1996).

To sum up, reasoning about social categories follows typical essentialist dimensions already in early childhood. Thus, our proposed structure of slurs neatly corresponds to the essentialist structure of social categories that is being tracked in cognitive psychology for the past 30 years. If we take these findings at face value, we must, in any case, grant that many social

---

[64] See also Prentice and Miller (2007) for an overview.

[65] This effect can't be accounted for by the explanation that children represent body features as more stable than non-body features. When 4-years olds were asked to choose whether body build vs. hair and skin color would remain the same over life span, they too judged significantly more often in favor of racial properties.

concepts have an essentialist structure. Thus, it is plausible and theoretically parsimonious to assume that also slur terms are associated with essentialistically structured concepts.

Let us now turn to the second key component of my semantics, namely, that slurs encode negatively valenced stereotypes. Recently, the phenomenon of slurring language has begun to be subjected to experimental testing by psycholinguists. These studies revealed that slurs, in contrast to their neutral counterpart terms, are *exclusively* associated with negative features. Because the authors of the studies used *implicit* paradigms in some of their studies, we have good reasons to think that these negative features belong to the *semantic* representation of slurs.[66]

In an experiment using a free association paradigm, Carnaghi and Maass (2008) delivered primary evidence for the negative stereotypes encoded in slurs. They presented participants with derogatory words (*fag*) or their neutral counterparts (*gay*). When presented with the slurs, the first three words participants mentioned were significantly more negatively-valenced than when presented with their neutral counterparts. However, since this experiment used an explicit paradigm, we cannot make strong inferences about the semantic structure of slurs on the basis of it.[67] For this reason, in a follow up study, Carnaghi and Maass used a *semantic priming task*. They presented participants *subliminally* with a prime word that was either a neutral term (*gay*), a derogatory counterpart term (*fag*), or a nonsense term (*secadftg*). Hence, the participants never consciously noticed with which word they were primed. This is important, as it eliminates the

---

[66] There are many ways to carve up the semantics-pragmatics distinction. In this paper, I assume the psychology-based framework according to which semantics includes those representations that enter into and are the result of immediate composition by our linguistic competence, and pragmatics includes all post-compositional representations that have been subject to general reasoning processing from central cognition.

[67] The most important limitation of explicit tasks is that they do not impose any constraints controlling for response modifications by conscious reasoning and voluntary control. For example, the negative association could as well be a post-semantic, pragmatic-inferential phenomenon and would thus not constitute evidence in favor of the hypothesis that stereotypes are semantically encoded in slurs. To reveal the 'bare' linguistic representations behind slurring words, it is more appropriate to employ a paradigm whose task outcomes are not influenced or distorted by other, non-semantic cognitive operations. Implicit tasks are ideal to unveil the semantic representations behind slurring words, since their task outcomes are less prone to be a result of intermixed high and low-level processes. This requirement was satisfied in Carnaghi and Maass' follow up study.

risk of task interventions by conscious higher-level pragmatic processes. Following the prime, the participants were to engage in a lexical decision task. They saw a target word that was either a trait stereotypical of the prime word (e.g., 'elegant' or 'effeminate'), counterstereotypical (e.g., 'energetic' or 'intolerant'), or completely unrelated ('honest' or 'stingy'). Importantly, half of the traits were positively valenced, while the other half of the target words was negatively valenced. The participants' task was to make a lexical word/nonword decision as fast as possible. The study had two key results. First, the participants reacted significantly faster to *stereotypical targets* than to counterstereotypical or unrelated targets, regardless of whether the prime was neutral ("gay") or derogatory ("fag"). This means that both neutral and derogatory category representations *immediately and automatically* activate representations of the related stereotype features. Second, the authors found that derogatory labels were again significantly less likely to activate flattering associations of the social group. As a matter of fact, derogatory labels resulted in the *suppression* of any positively-valenced stereotype, giving way for the negatively valenced associations related to a group.[68]

In sum, these experiments show that a) slur terms encode the stereotype associated with a social group,[69] and that b) this stereotype differs in valence from the stereotype encoded in the neutral counterpart, which confirms, experimentally, that there is a major semantic difference between neutral category labels and their corresponding epithets. In similar vein, also the authors of the studies conclude that

"[t]ogether, these results suggest that derogatory group labels differ from category group labels mainly with respect to the valence of the associations they elicit. Thus, it is not so much the ability to activate

---

[68] In a later study (Carnaghi & Maass, 2007), the authors successfully replicated the results, speaking for the robustness of their findings.

[69] Needless to say, this research, together with the abundant amount of research on typicality effects originating in Eleanor Rosch's research program (Rosch, 1988), supports that typicality effects are robust and thus relatively stable among subjects. Even critics of prototype theory often describe this stability as the most attractive feature of prototype theory (Fodor, 1998).

stereotypical content than the tendency to activate less-flattering associations that distinguishes derogatory from category group labels" (Carnaghi & Maass, 2007, p. 147).

After we have seen, in the last section, that my view accounts for the central linguistic data, I have now shown that my view receives additional, converging evidence from cognitive psychology. Specifically, we have seen that social concepts associated with social terms are organized essentialistically, and that slurs are uniquely associated with negatively valenced stereotypes. As a result, my semantic theory converges with an independently plausible research program about the essentialist structure of social concepts, and is directly supported by experimental research on slurring words.

3.2 Nomen est Omen: The Important Role of Nouns

If you go through your mental list of slurs that you are familiar with, you will probably notice that all of them belong to the syntactic category of nouns. This is peculiar, given that most of slurs' neutral counterparts come in both nominal and adjectival form (e.g., "a homosexual" / "homosexual"; "a Jew" / "Jewish"; notice also adverbial constructions such as "someone *with* homosexual preferences", "someone *from* Mexico", etc.). A linguistic theory of slurs should then be able to explain this systematic pattern of the data, and not treat it as a mere accident. Interestingly, this syntactic inflexibility of slurs is, too, a direct prediction of the essentialist theory. In contrast to other available theories, the essentialist theory is not only compatible with, but *makes sense* of the fact that nouns are the primary linguistic vehicle through which we communicate the semantic information of slurs.

There are a variety of different linguistic devices by means of which we can assign an individual to a category. Borrowing an example from Wierzbicka (1986), consider the difference between the sentences "Anna is blond" and "Anna is a blonde". In both cases, we predicate a

property (blondness) to Anna, and thereby include her in the set of things that instantiate 'blondness'. But although the set-theoretic operations of both predicates are *prima facie* identical, there are big disparities in the information conveyed by the adjective and the noun. Whereas the former predicate "blond" simply refers to a quality – a specific hair color – the latter predicate, "a blonde" is a sortal that refers to an object, or, rather, a person that can have a whole bunch of other qualities. Normally, we even feel compelled to make a number of inferences about *which* these other qualities are that Anna, being a blonde, has. By using the noun rather than the adjective, the speaker conveys that Anna is sexy or not particularly bright.

Another example, adopted from Gelman (2003), is a case in point. The Atlanta baseball player John Rocker was criticized for making a racist comment in an interview. When an ABC News reporter asked him directly, "Are you a racist?", he answered: "Absolutely not…. You hit one home run in the big leagues, it doesn't make you a home run hitter. … To make one [racist] comment like this doesn't make you a racist". Although Rocker's argument structure seems disputable (to say the least), it does tell us something about the underlying conceptual difference connected to a noun ("a homerun hitter" / "a racist") and a verb phrase ("to hit a home run" / "to make a racist comment"). Importantly, Rocker himself seems to take for granted that the verbal choice he makes directly conveys the difference in meaning between "to hit a home run" and to be "a homerun hitter". Whereas the first choice of syntactic category conveys a temporary state that does not originate in any identifying property of the person, the latter noun form ("a homerun hitter"/ "a racist") implicates an enduring, stable state that is central to the person's identity and reliably causes a number of other properties of the person (Gelman, 2003, p. 188). In short, nouns intuitively ("a racist", "a schizophrenic", "a blonde", "a liar", "a homerun hitter", etc.) impart a form of essentialism: the property that is nominalized is vital to the person's identity and allows for a variety of inductions. Other grammatical forms, such as adjectives and verb constructions ("have

schizophrenia" / "schizophrenic", "to have blond hair" / "to be blond", etc.) rather convey mutable, temporal qualities of an individual.

That nouns are much stronger in their essentialist-communicative potential than other word forms has received much empirical support. In one study, Susan Gelman and Gail Heyman compared the inductive potential children infer from noun and verb labels (Gelman & Heyman, 1999). They either heard a story that contained "a carrot eater" (noun phrase; NP), or a story that talked about someone who "eats carrots whenever she can" (verb phrase; VP). In the critical part, the children answered a set of questions that tested the stability of the properties: e.g., "Will Rose eat a lot of carrots when she is grown up", or "Would Rose eat a lot of carrots if she grew up in a family where no one liked carrots?" Children in the NP condition predicted significantly more often that the property in question would be more stable over time and in adverse environmental conditions than children in the VP condition. Thus, the grammatical form of a noun seems to suggest to a child that a category is to be thought of as a kind (Gelman & Heyman, 1999).

Carnaghi et al. (2008) replicated and developed the experiments initiated by Gelman and Heyman. In six experiments testing adults, they compared the inductive potential of nouns and adjectives which assign individuals to the same categories (e.g., "an athlete" vs. "athletic"). They found that describing a person by a noun triggers significantly more stereotypical inferences as compared to an adjectival description. Remarkably, nouns also *inhibit* inferences about behaviors or habits that are associatively rather incongruent with the descriptors. For example, a person that is homosexual (adjective) was estimated to attend the church more often than *a* homosexual (noun). Moreover, nouns but not adjectives inhibit the possibility of alternative classifications *altogether* (i.e., not only incongruent ones). Once someone is categorized as belonging to one social category, e.g., "artist", participants are not very willing to assign them to a second one, e.g., "athlete". These results did not hold for adjective conditions, because nouns as opposed to

adjectives tend to convey discrete category boundaries which do not intersect with other categories. Finally, when Carnaghi et al. primed subjects with an essentialist scenario, participants would even *themselves* be more likely to use a noun to describe a person.

In sum, nouns, adjectives and verb phrases do not only *categorize* individuals, but also tell us something about the particular *way* in which the individuals are categorized. In the case of adjectives, the individual is assigned to one qualitative category among many potential others. In the case of nouns, the individual is assigned to one category that *identifies* the individual in question in a rather all-or-nothing way and allows for rich inferences with regard to qualities that (allegedly) come along with the stable category in question.

It is clear how the essentialist theory explains that nouns are the main syntactic vehicle of slurs. According to my theory, slurs essentially encode essentialist information. We have now seen that nouns are the primary linguistic device we use to convey that a category is essentialized. On my account, then, nouns *should* be the linguistic tool to communicate the meaning of slurs. In sum, then, nouns 'efficiently' communicate the information encoded in slurs and thus play a crucial role for the formation and propagation of slurs.

Conclusion

The closing scene of *I Am Not Your Negro* shows James Baldwin issuing a powerful, penetrating diagnosis of White America.

"What white people have to do, is try to find out in their hearts why it was necessary for them to have a nigger in the first place. Because I am not a nigger. I'm a man. If I'm not the nigger here, and if you invented him, you the white people invented him, then you have to find out why. And the future of the country depends on that. Whether or not it is able to ask that question." (Baldwin, 2016/1963)[70]

---

[70] Raoul Peck's documentary film *I Am Not Your Negro* is a collage based on the unfinished manuscripts of *Remember this House*, immersed with interview excerpts by Baldwin, and a variety of other material (Peck, 2016). The final scene

This paper was an attempt to contribute to the task assigned so poignantly by Baldwin, and come a small step closer to an answer to his question. I have argued that the central mechanism of slurs is one of *essentialization*; slurs are akin to kind terms that denote an essence of a social category which nomologically connects to a set of negative stereotypical features. In effect, by using slurs, or even having them in our public lexicon, we commit to a way of carving up the social world in a way that is dehumanizing and gives groups with the dominant share of social power a tool to rationalize and maintain the oppressive hierarchies that keep down marginalized groups.

To illustrate the plausibility of the advanced hypothesis, and show that it does interesting, multi-layered explanatory work, I argued that, first, essentialism about slurs explains their recognized linguistic properties, second, that the essentialist theory receives convergent evidence from cognitive psychology, and third, that the essentialist theory has unique resources to explain why slurs occur predominantly as nouns. Importantly, the goal of this paper has been to make a *cumulative* case for the essentialist theory and motivate it as a novel, interesting framework that takes seriously the challenge of linking racist language to cognition and explains its relation to social oppression. Although someone might disagree with the assessment of some of the data, 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 this paper illustrated that the essentialist theory does precisely that.

Bibliography

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

---

the quote is based on is originally from a 1963 interview of Baldwin with Kenneth Clar. Note that the invention of the "nigger" by the white world is a re-occurring theme employed by Baldwin (see, for example, Baldwin, 1969).

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

Appiah, K. A. (2018). *The Lies That Bind: Rethinking Identity*. Liveright.

Bach, K. (2018). Loaded words: On the semantics and pragmatics of slurs. In D. Sosa (Ed.), *Bad Words*. Oxford University Press.

Baldwin, J. (1969). THE NIGGER WE INVENT. *Equity & Excellence in Education*, *7*(2), 15–23. https://doi.org/10.1080/0020486690070203

Basu, R. (2019). What We Epistemically Owe To Each Other. *Philosophical Studies*, 1–17. https://doi.org/10.1007/s11098-018-1219-z

Bero, S. (2017). Responsibility and the Emotional Structure of Relationships. USC Dissertation.

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. https://doi.org/10.1521/soco.2000.18.4.329

Block, N. (1987). Advertisement for a Semantics for Psychology. *Midwest Studies in Philosophy*, *10*(1), 615–678. https://doi.org/10.1111/j.1475-4975.1987.tb00558.x

Bolinger, R. J. (2017). The Pragmatics of Slurs. *Noûs*, *51*(3), 439–462. https://doi.org/10.1111/nous.12090

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

Camp, E. (2018). A Dual Act Analysis of Slurs. In D. Sosa (Ed.), *Bad Words: Philosophical Perspectives on Slurs* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780198758655.003.0003

Carnaghi, A., & Maass, A. (2007). In-Group and Out-Group Perspectives in the Use of Derogatory Group Labels: Gay Versus Fag. *Journal of Language and Social Psychology*, *26*(2), 142–156. https://doi.org/10.1177/0261927X07300077

Carnaghi, A., & Maass, A. (2008). Derogatory language in intergroup context: Are "gay" and "fag" synonymous? In Y. Kashima, K. Fiedler, & P. Freytag (Eds.), *Stereotype dynamics: Language-based approaches to the formation, maintenance, and transformation of stereotypes*. (pp. 117–134).

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

Cepollaro, B. (2017). *The Semantics and Pragmatics of slurs and thick terms*. PSL Research University.

Cepollaro, B., & Stojanovic, I. (2016). Hybrid Evaluatives: In Defense of a Presuppositional Account. *Grazer Philosophische Studien*, *93*(3), 458–488. https://doi.org/10.1163/18756735-09303007

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press.

Copp, D. (2009). Realist-expressivism and conventional implicature. *Oxford Studies in Metaethics*, *4*, 167–202.

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

Croom, A. M. (2011). Slurs. *Language Sciences, 33*(3), 343–358. https://doi.org/10.1016/J.LANGSCI.2010.11.005

Croom, A. M. (2014a). Remarks on "The Semantics of Racial Slurs." *Linguistic and Philosophical Investigations, 13*, 11–32.

Croom, A. M. (2014b). Spanish slurs and stereotypes for Mexican-Americans in the USA: A context-sensitive account of derogation and appropriation. *Pragmática Sociocultural / Sociocultural Pragmatics, 8*(2), 145. https://doi.org/10.1515/soprag-2014-0007

Croom, A. M. (2015). The semantics of slurs: A refutation of coreferentialism. *Ampersand, 2*, 30–38. https://doi.org/10.1016/J.AMPER.2015.01.001

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

Danks, D. (2014). *Unifying the mind: cognitive representations as graphical models.* MIT Press.

Del Pinal, G. (2016). Prototypes as compositional components of concepts. *Synthese, 193*(9), 2899–2927. https://doi.org/10.1007/s11229-015-0892-0

Del Pinal, G. (2018). Meaning, modulation, and context: a multidimensional semantics for truth-conditional pragmatics. *Linguistics and Philosophy, 41*(2), 165–207. https://doi.org/10.1007/s10988-017-9221-z

Demoulin, S., Leyens, J.-P., & Yzerbyt, V. (2006). Lay Theories of Essentialism. *Group Processes & Intergroup Relations, 9*(1), 25–42. https://doi.org/10.1177/1368430206059856

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

Diesing, M. (1992). *Indefinites.* MIT Press.

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

Dovidio, J. F., Glick, P. S., & Rudman, L. A. (Eds.). (2005). *On the nature of prejudice: Fifty years after Allport.* Malden, MA: Blackwell Pub. https://doi.org/10.1002/9780470773963

Fodor, J. (1998). *Concepts: Where cognitive science went wrong.* Oxford and New York: Oxford University

Press.

Fodor, J. (2005). *Hume Variations.* Oxford; Oxford; New York: Clarendon Press. https://doi.org/10.1093/acprof:oso/9780199287338.001.0001

Fodor, J., & Lepore, E. (1992). *Holism. A shopper's guide.* Blackwell.

Fodor, J., & Pylyshyn, Z. W. (2014). *Minds without meanings: an essay on the content of concepts.* MIT Press.

Gädenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought.* MIT Press.

Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46*(1), 23. https://doi.org/10.2307/3033657

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

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences, 8*(9), 404–409. https://doi.org/10.1016/j.tics.2004.07.001

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring Properties from Categories versus Inferring Categories from Properties: The Case of Gender. *Child Development, 57*(2), 396–404. https://doi.org/10.2307/1130595

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. https://doi.org/10.1111/1467-9280.00194

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. J. (2001). Are Ethnic Groups Biological "Species" to the Human Brain? *Current Anthropology, 42*(4), 515–553. https://doi.org/10.1086/321802

Harman, G. (1993). Holism: A Consumer Update. *Grazer Philosophische Studien, 46,* 163–171. https://doi.org/10.5840/gps1993467

Haslam, N. (2000). Psychiatric categories as natural kinds: Essentialist thinking about mental disorder. *Social Research, 67*(4), 1031–1058.

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 and 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. *The British Journal of Social Psychology, 39 (Pt 1)*(1), 113–127. https://doi.org/10.1348/014466600164363

Haslam, N., Rothschild, L., Ernst, D., Yzerbyt, V., Judd, C. M., & Corneille, O. (2004). Essentialism and Entitativity: Structures of Beliefs about the Ontology of Social Categories. *The Psychology of Group Perception: Perceived Variability, Entitativity, and Essentialism.,* 61–78.

https://doi.org/10.4324/9780203644973

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar* (Vol. 13). Blackwell Oxford.

Hirschfeld, L. A. (1995). Do children have a theory of race? *Cognition, 54*(2), 209–252. https://doi.org/10.1016/0010-0277(95)91425-R

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

Ho, A. K., Roberts, S. O., & Gelman, S. A. (2015). Essentialism and Racial Bias Jointly Contribute to the Categorization of Multiracial Individuals. *Psychological Science, 26*(10), 1639–1645. https://doi.org/10.1177/0956797615596436

Hom, C. (2008). The Semantics of Racial Epithets. *Journal of Philosophy, 105*(8), 416–440. https://doi.org/10.5840/jphil2008105834

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

Hom, C., & May, R. (2013). Moral and Semantic Innocence. *Analytic Philosophy, 54*(3), 293–313. https://doi.org/10.1111/phib.12020

Hom, C., & May, R. (2015). Pejoratives as Fiction. In D. Sosa (Ed.), *Bad Words* (pp. 1–22). Oxford University Press.

Hornsby, J. (2001). Meaning and Uselessness: How to Think about Derogatory Words. *Midwest Studies in Philosophy, 25*(1), 128–141. https://doi.org/10.1111/1475-4975.00042

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. https://doi.org/10.1111/phpe.12027

Jeshion, R. (2013b). Slurs and Stereotypes. *Analytic Philosophy, 54*(3), 314–329. https://doi.org/10.1111/phib.12021

Jeshion, R. (2018). Slurs, Dehumanization, and the Expression of Contempt. In D. Sosa (Ed.), *Bad Words* (Vol. 1). Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198758655.003.0005

Jönsson, M. L. (2017). Interpersonal Sameness of Meaning for Inferential Role Semantics. *Journal of Philosophical Logic, 46*(3), 269–297. https://doi.org/10.1007/s10992-016-9400-3

Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition, 57*(2), 129–191. https://doi.org/10.1016/0010-0277(94)00659-9

Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition, 127*(2), 242–257. https://doi.org/10.1016/j.cognition.2013.01.005

Lakoff, G. (1987). *Women, fire, and dangerous things: what categories reveal about the mind.* University of Chicago Press.

Leslie, S.-J. (2013). *Oxford Studies in Epistemology Volume 4.* (T. S. Gendler & J. Hawthorne, Eds.), *Oxford studies in epistemology* (Vol. 4). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199672707.001.0001

Leslie, S.-J. (2015). "Hillary Clinton is the Only Man in the Obama Administration": Dual Character Concepts, Generics, and Gender. *Analytic Philosophy, 56*(2), 111–141. https://doi.org/10.1111/phib.12063

Leslie, S.-J. (2017). The Original Sin of Cognition. *Journal of Philosophy, 114*(8), 393–421. https://doi.org/10.5840/jphil2017114828

Livingstone Smith, D. (2011). *Less Than Human: Why We Demean, Enslave, and Exterminate Others.* St. Martin's Press.

Mahalingam, R. (2003). Essentialism, Culture, and Power: Representations of Social Class. *Journal of Social Issues, 59*(4), 733–749. https://doi.org/10.1046/j.0022-4537.2003.00087.x

Marconi, D. (1997). *Lexical Competence.* MIT Press.

McCawley, J. D. (1972). A Program for Logic. In *Semantics of Natural Language* (pp. 498–544). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-2557-7_18

Medin, D., & Ortony, A. (1989). Psychological essentialism. *Similarity and Analogical Reasoning, 179,* 179–195. https://doi.org/10.1017/cbo9780511529863.009

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* (pp. 192–207). Malden, MA: Blackwell Pub. https://doi.org/10.1002/9780470773963.ch12

Murphy, G. L. (2004). *The big book of concepts.* Cambridge, Mass.: MIT Press.

Nunberg, G. (2018). The Social Life of Slurs. In D. Fogal, D. W. Harris, & M. Moss (Eds.), *New Work on Speech Acts.* Oxford University Press. https://doi.org/10.1093/oso/9780198738831.003.0010

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

Peck, R. (2016). *I am not your Negro.* United States: Velvet Film, Inc.

Popa-Wyatt, M. (2016). Not All Slurs are Equal. *Phenomenology and Mind, 11,* 150–156. https://doi.org/http://dx.doi.org/10.13128/Phe_Mi-20115

Potts, C. (2004). *The Logic of Conventional Implicatures.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199273829.001.0001

Potts, C. (2007). The expressive dimension. *Theoretical Linguistics, 33*(2), 165–198. https://doi.org/10.1515/TL.2007.011

Prentice, D. A., & Miller, D. T. (2006). Essentializing differences between women and men. *Psychological Science, 17*(2), 129–135. https://doi.org/10.1111/j.1467-9280.2006.01675.x

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

Prinz, J. J. (2012). Regaining Composure: A Defence Of Prototype Compositionality. In W. Hinzen, M. Werning, & E. Machery (Eds.), *The Oxford Handbook of Compositionality*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199541072.013.0021

Putnam, H. (1975). The meaning of "meaning." *Minnesota Studies in the Philosophy of Science, 7*, 131–193.

Rehder, B. (2017). Concepts as Causal Models. In M. R. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.39

Rey, G. (1983). Concepts and stereotypes. *Cognition, 15*(1–3), 237–262. https://doi.org/10.1016/0010-0277(83)90044-6

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

Richard, M. (2008). *When Truth Gives Out*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199239955.001.0001

Rosch, E. (1988). Principles of Categorization. In *Readings in Cognitive Science* (pp. 312–322). Elsevier. https://doi.org/10.1016/B978-1-4832-1446-7.50028-5

Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? *Language, Interaction and Social Cognition*, 11–36. https://doi.org/10.1002/casp.2450040509

Schlenker, P. (2007). Expressive presuppositions. *Theoretical Linguistics, 33*(2), 237–245. https://doi.org/10.1515/TL.2007.017

Sennet, A., & Copp, D. (2015). What kind of a mistake is it to use a slur? *Philosophical Studies, 172*(4), 1079–1104. https://doi.org/10.1007/s11098-014-0338-4

Sloman, S. (2005). *Causal Models*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195183115.001.0001

Sloman, S., Love, B. C., & Ahn, W.-K. (1998). Feature Centrality and Conceptual Coherence. *Cognitive Science, 22*(2), 189–228. https://doi.org/10.1207/s15516709cog2202_2

Smith, E. E., Medin, D. L., & Rips, L. J. (1984). A psychological approach to concepts: Comments on Rey's "Concepts and stereotypes." *Cognition, 17*(3), 265–274. https://doi.org/10.1016/0010-0277(84)90010-6

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

Soames, S. (2015). *Rethinking language, mind, and meaning*. Princeton University Press.

Strawson, P. F. (2012). *Introduction to logical theory*. Routledge.

Taylor, J. R. (2009). *Linguistic categorization* (3. ed., re). Oxford [u.a.]: Oxford Univ. Press.

Verkuyten, M. (2003). Discourses about ethnic group (de-)essentialism: Oppressive and progressive

aspects. *British Journal of Social Psychology, 42*(3), 371–391. https://doi.org/10.1348/014466603322438215

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

Whiting, D. (2013). It's Not What You Said, It's the Way You Said It: Slurs and Conventional Implicatures. *Analytic Philosophy, 54*(3), 364–377. https://doi.org/10.1111/phib.12024

Wierzbicka, A. (1986). What's in a Noun? (Or: How Do Nouns Differ in Meaning from Adjectives?). *Studies in Language, 10*(2), 353–389. https://doi.org/10.1075/sl.10.2.05wie

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