

Special Issue

## Value in Language

Dan Zeman

Guest Editor

### Contents

Dan Zeman: *Introduction: "Value in Language"* ..... 498

#### Research Articles

Pekka Väyrynen: *Normative Naturalism on Its Own Terms* ..... 505

Natalia Karczewska: *Illocutionary Disagreement in Faultless  
Disagreement* ..... 531

Alex Davies: *Faultless Disagreement Contextualism* ..... 557

Katharina Felka: *'Boys Don't Cry' – An Ambiguous Statement?* ..... 581

Andrés Soria-Ruiz: *Value and Scale: Some Observations and  
a Proposal* ..... 596

Chang Liu: *The Derogatory Force and the Offensiveness of Slurs* ..... 626

Alice Damirjian: *Rethinking Slurs: A Case Against Neutral Counterparts  
and the Introduction of Referential Flexibility* ..... 650

Bianca Cepollaro: *The Moral Status of the Reclamation of Slurs* ..... 672

Zuzanna Jusińska: *Slur Reclamation – Polysemy, Echo, or Both?* ..... 689

Alba Moreno – Eduardo Pérez-Navarro: *Beyond the Conversation:  
The Pervasive Danger of Slurs* ..... 708

Stefano Predelli: *Unmentionables: Some Remarks on Taboo* ..... 726

## Introduction: “Value in Language”

Valuing seems to be a fundamental human endeavor. We constantly attribute positive and negative traits to people, actions, events and objects around us. Thus, we find a character trait admirable, an action commendable, a dish delicious, a piece of music beautiful, a reasoning correct and so on—as well as the opposite. As we move in the world, we are guided by our evaluations, by seeking what is valuable and avoiding what is not.

We also talk a lot about valuing and about what is valuable, and natural languages have plenty of expressions that allow us to do that, in more direct or indirect ways. Given our many worldviews and goals, we also quite often disagree about what is or is not valuable, as well as about how to express it. How exactly to connect value and valuing to the meaning of the expressions we use to do all that is a question that has been on philosophers of language’s minds for a long time. It is, still, an open question.

This special issue hosts 11 papers that tackle various questions that arise in relation to value and valuing

in language. The interests of the authors featured range from general considerations regarding the normative sphere to very specific issues and phenomena connected to the use of various natural language expressions such as predicates of taste, generics, evaluatives, slurs and taboo words. This introduction serves to present the special issue to the reader by giving a short description of the main claims and arguments of each contribution.

The issue opens with Pekka Väyrynen’s paper “Normative Naturalism on Its Own Terms”. In it, Väyrynen focuses on normative naturalism (the thesis that normative facts and properties are among natural facts and properties) and investigates two claims related to how we talk and think about norms. The first is that a successful naturalist view requires describing normative properties in wholly non-normative terms. Using arguments found in the literature (especially those by Sturgeon (2003)) and offering some of his own, Väyrynen argues that providing the said description is not a commitment of normative



naturalism *per se* and that this does not threaten the notion of natural property. The second claim is that normative properties are "too different" from natural properties to be counted as such. Here, Väyrynen shows, first, that once the previous point is acknowledged, the objection loses its force; and second, that the notion of "genuine autoritativity" (what makes normative concepts play the special role they do: their connection to deliberation, decision, action etc.) is hard to pin down. The paper also contains a "proof of concept" (in terms of orderings of non-arbitrary selections among the live options in a given deliberative context) that normative naturalists have no troubles accounting for thought and talk related to genuine autoritativity.

The next two papers tackle the much-discussed phenomenon of faultless disagreement, said to be found chiefly in "disputes of inclination"—for example, about matters of taste. Natalia Karczevska's paper investigates this phenomenon in relation to the predicate of personal taste 'tasty'. Thus, in "Illocutionary Disagreement in Faultless Disagreement", she argues that extant contextualist proposals fail to account for autocentric disagreements involving 'tasty' and proposes a novel view of disagreement that does. Karczevska takes predicates of taste to be associated with

a new type of illocutionary speech act—what she calls "evaluations", which she thoroughly characterizes following Searle and Vanderveken's (1985) list of features. Disagreement arises due to failed attempts at introducing opposite commitments imposed on the common ground by such acts. Karczevska further argues that this way of seeing disagreement, although close to the more familiar "clash of attitudes" construal (originating with Stevenson 1963), is nevertheless different from it and less troublesome.

Alex Davies also picks up the issue of faultless disagreement in his contribution, but takes the discussion in a different direction. Davies argues in "Faultless Disagreement Contextualism" that whether a certain exchange is a faultless disagreement essentially depends on context. His main target is the widely assumed idea that the source of the phenomenon is the meaning of the target predicates themselves—that is, their "subjective" character, which distinguishes them from "objective" ones. By carefully going through a wide range of examples and by putting forward a positive proposal that connects faultless disagreement with the reasons interlocutors have for making their assertions (so that it arises when those reasons are permissive with respect to assessing whether

a certain object has a property), Davies shows that this assumption doesn't hold. One (important) consequence of this view is that the metasemantics of "subjective" predicates is context-sensitive, thus undercutting the debate between contextualism and relativism. The final part of the paper is dedicated to answering four objections to this way of seeing faultless disagreement.

The next paper in the issue—Katharina Felka's "‘Boys Don't Cry’ - An Ambiguous Statement?"—focuses on generic statements. Specifically, Felka aims to show that sentences like the one in her title that have what is called a "normative" reading should not be given a semantic treatment, but a pragmatic one instead. Felka proposes that such readings are best taken to be conversational implicatures, generated by the maxim of relation ('Be relevant!'). She engages with two prominent views on generics, Leslie's (2015) and Cohen's (2001), showing their inadequacy to capture normative readings, and arguing at the same time that a pragmatic account like the one described above has all the resources needed to do so.

'Good' is one of the English words perhaps most closely connected with valuing and value. In "Value and Scale: Some Observations and a Proposal", Andrés Soria-Ruiz sets out to disentangle what semantic treatment

is best suited for it. Starting from the observation that 'good' is gradable, Soria-Ruiz investigates what type of scale should be associated with the word, and argues that a novel type—"round ratio scales"—is the answer. In doing so, he operates within the framework proposed by Lassiter (2017), but enriches and transforms it so that to accommodate various linguistics phenomena (most importantly the felicity of expressions like 'twice as good') that Lassiter's framework in itself was not able to. One notable consequence of Soria-Ruiz's view is that there is a rift between propositional level and individual level 'good': while the former has an interval scale, round ratio scales apply to the latter.

A slew of papers in this issue are concerned with slurs—proving once again how attractive for researchers this topic has been in recent years. While the range of topics dealt with varies from general or more fundamental issues to very specific ones, all papers contribute to the elucidation of some important aspect of the current, multifaceted debate involving slurs. Thus, in "The Derogatory Force and the Offensiveness of Slurs", Chang Liu argues for the importance of clearly distinguishing between the two elements mentioned in his title and that neglecting this distinction in current literature has led to muddling the

waters. Four arguments are presented: from a comparison with non-slurs, from the behavior of quoted slurs, from the use of slurs in argots and from the difference between derogatory and offensive autonomy. He also puts forward a positive view ("The Speech Act Theory of Slurs") according to which derogation and offence are achieved via specific speech acts (the former illocutionary, the latter perlocutionary) and compares this view with a few others on the market (Anderson and Lepore's (2013) prohibition view; Davis and McCready's (2020) invocational view; etc.), showcasing its advantages.

With her paper "Rethinking Slurs: A Case Against Neutral Counterparts and the Introduction of Referential Flexibility", Alice Damirjian brings into discussion an idea that has played a big role in the debate so far: namely, that slurs have what is known as "neutral counterparts" (expressions that refer to the same group as a slur but don't contain an evaluative component). Focusing on Diaz Legaspe's (2018) defense of this claim, Damirjian forcefully opposes it by adducing arguments both from past and present uses of slurs in support of the idea of "referential flexibility": the fact that slurs are often used to refer to a subgroup of members of their presumed neutral counterparts, but also to individuals that don't belong to the group.

This shows that slurs and their presumed neutral counterparts cannot be truth-conditionally equivalent. She maintains not only that the neutral counterparts idea is unsupported by the data, but also that assuming it in current debates leads us astray in our inquiry.

Bianca Cepollaro's paper "The Moral Status of the Reclamation of Slurs" concerns reclamation: the act of taking a negatively-charged expression such as a slur and turning it into a positive one for political, solidarity or camaraderie purposes. Specifically, Cepollaro engages with an argument against the legitimacy of reclamation ("the warrant argument"): namely, that since no negative evaluative property can be essentially connected to a non-evaluative one, neither a positive one should. This puts reclamation into doubt. Cepollaro carefully spells out the premises of the argument and then replies to it by making a parallel with affirmative action: as it can be morally permissible to balance an existing form of injustice by introducing a mechanism that temporarily violates the relevant norm of equality, so reclamation can be morally permissible too. The paper ends with some remarks aimed at debunking "the myth of reverse racism and sexism".

Zuzanna Jusińska is concerned with the same phenomenon, albeit with a different purpose in mind. In

"Slur Reclamation—Polysemy, Echo, or Both?", Jusińska is interested in the precise mechanism by which reclamation works, and to this effect they investigate two prominent views on the market: Jeshion's (2020) and Bianchi's (2014). Jusińska takes each of them not to be wrong, but incomplete, and thus to support the need for a more complex picture. Jusińska appeals to detailed historical records of the reclamation of certain slurs ('queer' and the n-word), which they take to mandate the introduction of an additional pragmatic step in Jeshion's scheme, involving echoic uses of slurs, that leads to the initiation of a new linguistic convention. The result of this endeavor is what is called in the paper "the Combined view of reclamation", which they take both to provide the element missing in previous accounts and to better handle the historical data.

The part of the issue tackling slurs ends with Alba Moreno and Eduardo Pérez-Navarro's paper "Beyond the Conversation: The Pervasive Danger of Slurs". The authors defend the view that occurrences of slurs, both in speech and in written form (including in academic papers) are dangerous in that they have the potential to be harmful. First, they reject the idea that whether a slur is derogatory depends on the linguistic environment it appears in: quoted slurs, for example,

can harm too. Second, they pin the derogatoriness of slurs on the type of context in which they occur: while slurs are always derogatory in "uncontrolled contexts", they can be non-derogatory in "controlled contexts" (roughly, the ones in which speakers know how they will be interpreted). However, their claim is that even the use of slurs in such contexts can lead to normalizing derogation. The authors end the paper with some considerations relevant for the practice of researchers writing on slurs.

The closing paper of the issue tackles a puzzling phenomenon that has interested scholars from various fields of inquiry: taboo. In "Unmentionables: Some Remarks on Taboo", Stefano Predelli ponders on what makes taboo words puzzling and shows that neither an orthodox, truth-conditional approach nor a more sophisticated, non-truth-conditional treatment fully accounts for it. On the positive side, Predelli gestures towards a theory that subsumes taboo words under a theory of action, as they are essentially related to acts of tokening, which neatly ties the fact that they are unmentionable to their mere occurrence in speech or writing. The last part of the paper contains some remarks about the silencing power of taboo and a plea for the fruitfulness of inquiring about taboo words—in itself, for the semantics

and for the metasemantics of natural language.

Both the topics of value and valuing in themselves and the range of specific issues related to them go far beyond what has been addressed in these papers. However, by putting together this special issue, I hope to have offered the reader a snapshot of some of the current preoccupations with these topics and issues and, hopefully, a springboard for future developments, arguments, and discussion. Obviously, the special issue wouldn't have been possible without the effort of a great number of people involved. Thus, I want to thank all contributors for their papers and their commitment to improve them, all the reviewers for their patience

and dedication, all those involved in the activities leading to the publication of the issue (especially Matteo Pascucci and Mirco Sambrotta for co-organizing with me the “Value in Language” workshop at the Slovak Academy of Sciences on 29-30.03.2021, on which the volume is based), and the Slovak Academic Information Agency for financial support (through the Initiative Project no. 2019-10-15-007). Last but not least, I want to thank the editor-in-chief of *Organon F*, Martin Vacek, for his unflinching support for this project.

Dan Zeman

University of Warsaw  
danczeman@gmail.com

## References

- Anderson, Luvell, and Ernie Lepore. 2013. “What Did You Call Me? Slurs as Prohibited Words.” *Analytic Philosophy* 54 (3): 350–63. <https://doi.org/10.1111/phib.12023>
- Bianchi, Claudia. 2014. “Slurs and Appropriation: An Echoic Account.” *Journal of Pragmatics* (66): 35–44. <https://doi.org/10.1016/j.pragma.2014.02.009>
- Cohen, Ariel. 2001. “On the Generic Use of Indefinite Singulars.” *Journal of Semantics* 18 (3): 183–209. <https://doi.org/10.1093/jos/18.3.183>
- Davis, Christopher, and Elin McCready. 2020. “The Instability of Slurs.” *Grazer Philosophische Studien* 97 (1): 63–85. <https://doi.org/10.1163/18756735-09701005>
- Diaz Legaspe, Justina. 2018. “Normalizing Slurs and Out-Group Slurs: The Case of Referential Restriction.” *Analytic Philosophy* 59 (2): 234–55. <https://doi.org/10.1111/phib.12129>
- Jeshion, Robin. 2020. “Pride and Prejudiced: on the Reclamation of Slurs.” *Grazer Philosophische Studien* 97 (1): 106–37. <https://doi.org/10.1163/18756735-09701007>
- Lassiter, Daniel. 2017. *Graded Modality: Qualitative and Quantitative Perspectives*. Oxford: Oxford University Press.

- 
- Leslie, Sarah-Jane. 2015. "‘Hillary Clinton is the Only Man in the Obama Administration’: Dual Character Concepts, Generics, and Gender." *Analytic Philosophy* 56 (2): 111–41. <https://doi.org/10.1111/phib.12063>
- Searle, John and Daniel Vanderveken. 1985. "Speech Acts and Illocutionary Logic." In *Logic, Thought and Action. Logic, Epistemology, and the Unity of Science* (volume 2), edited by Daniel Vanderveken, 109–32. Springer. [https://doi.org/10.1007/1-4020-3167-X\\_5](https://doi.org/10.1007/1-4020-3167-X_5)
- Stevenson, Charles. 1963. *Facts and Values: Studies in Ethical Analysis*. New Haven, CT: Yale University Press.
- Sturgeon, Nicholas L. 2003. "Moore on Ethical Naturalism." *Ethics* 113 (3): 528–56. <https://doi.org/10.1086/345627>



## Normative Naturalism on Its Own Terms

Pekka Väyrynen\*

Received: 6 November 2020 / Accepted: 9 April 2021


*Abstract:* Normative naturalism is primarily a metaphysical doctrine: there are normative facts and properties, and these fall into the class of natural facts and properties. Many objections to naturalism rely on additional assumptions about language or thought, but often without adequate consideration of just how normative properties would have to figure in our thought and talk if naturalism were true. In the first part of the paper, I explain why naturalists needn't think that normative properties can be represented or ascribed in wholly non-normative terms. If so, certain prominent objections to normative naturalism fail. In the second part, I consider the objection that normative properties are “just too different” from (other) natural properties to themselves be natural properties. I argue that naturalists have no distinctive trouble making sense of thought and talk involving forms of “genuine” or “authoritative” normativity which can drive a non-question-begging form of the objection.

*Keywords:* Authoritative normativity; normative concepts; normative naturalism; one-term naturalism; practical normativity.

---

\* University of Leeds

 <https://orcid.org/0000-0003-4066-8577>

 The School of Philosophy, Religion and History of Science, University of Leeds, Leeds, LS2 9JT, United Kingdom.

 [p.vayrynen@leeds.ac.uk](mailto:p.vayrynen@leeds.ac.uk)

---

© The Author. Journal compilation © The Editorial Board, *Organon F*.



This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International Public License (CC BY-NC 4.0).

## 1. Introduction

Normative naturalism is, to a first approximation, the view that there are normative facts and properties, and these fall into the class of natural facts and properties.<sup>1</sup> Specific forms of naturalism may come with additional semantic, epistemological, or other commitments. At its core, however, normative naturalism is a metaphysical doctrine. But many objections to naturalism rely on additional assumptions about language or thought. My aim in this paper is to make two (largely distinct) contributions to debates about how normative properties might figure in language and thought if normative naturalism is true.

The first part of the paper focuses on an assumption I'll call Non-Normative Representability (NNR). To a first approximation, NNR says that nothing counts as a natural property unless it can be expressed, or represented, or ascribed with wholly non-normative terms or concepts. Paradigmatic non-normative terms include 'is tubular', 'has low air pressure', and 'promotes survival'. The upshot of NNR for normative properties is that their being natural depends on whether they can be ascribed not only by normative terms or concepts, such as 'ought', 'wrong', and 'good', but also by non-normative terms or concepts. Many naturalists accept this. But Nicholas Sturgeon (2003) has argued that NNR isn't a commitment of normative naturalism as such. The point is worth laboring because it has important ramifications, but it keeps getting ignored. I'll improve on Sturgeon's statement of NNR a bit and illustrate what's at stake by explaining how a wide range of objections to normative naturalism presuppose NNR (§2). I'll then offer reasons, Sturgeon's and my own, why the truth of normative naturalism doesn't require NNR and why NNR is questionable enough for naturalists to have reason to keep their distance since they can (§3). I'll also discuss why this needn't mean losing our grip on the notion of a natural property (§4).

---

<sup>1</sup> I use 'normative' to cover both the deontic and the evaluative. I'll understand properties as entities that characterize the objects which have them. I use 'property' broadly to cover also relations. I'll take a fact to be an entity, a state of affairs, which concerns objects exemplifying properties or standing in relations.

The second part of the paper offers a slightly sideways approach to the “just too different” objection to normative naturalism. Many critics of naturalism think that especially properties involving “robust” or “authoritative” normativity are too different from (other) natural properties to also be natural properties. I’ll first suggest that the objection loses some of its force if normative naturalism isn’t committed to NNR and note difficulties in specifying the notion of “genuine” or “authoritative” normativity which is the objection’s primary concern (§5). Some notions of such authority are too weak to support the objection, but many stronger notions are question-begging. I’ll then pick a particular authoritatively normative concept which falls somewhere in the middle to give a proof of concept that naturalists have no distinctive trouble making sense of thought and talk involving authoritative normativity (§6). This strategy doesn’t require rejecting NNR, but avoids some headaches without it. Thus, I’ll focus throughout on forms of normative naturalism which needn’t accept NNR.<sup>2</sup>

## 2. Non-normative representability and objections to naturalism

To get a better grip on Non-Normative Representability and what’s at stake in it, it’s instructive to consider some objections to normative naturalism which presuppose NNR.

The observation that discussions of normative naturalism often presuppose NNR isn’t original to me. In its standard interpretations, G. E. Moore’s “open question” argument implies that ‘good’ doesn’t stand for a natural property, roughly on the grounds that it cannot be analyzed or defined in any wholly non-normative terms (Moore 1903, ch. 1). In a rich discussion of Moore’s arguments against naturalism, Sturgeon notes that we know from the beginning that ‘good’ is coreferential with itself. It’s only if you assume from the outset that ‘good’ doesn’t stand for a natural property in its own right that an argument that ‘good’ is indefinable shows that ‘good’

---

<sup>2</sup> I won’t consider “analytic” naturalism. This implies NNR, since it says that any normative predicate is analytically equivalent with, and in principle replaceable by, a descriptive, non-normative predicate that ascribes the same property. For a sophisticated contemporary form of this view, see (Jackson 1998).

isn't coreferential with any term whatever standing for a natural property.<sup>3</sup> Sturgeon offers a conjecture regarding why generations of critics have missed that the open question argument begs the question in this way: "I think that the answer must be that they are relying on an assumption about natural properties that seems to them so obvious as not to need stating: namely, that nothing counts as a natural property unless we have some non-ethical terminology to represent it" (Sturgeon 2003, 536). Sturgeon here states a restricted thesis. If we generalize his talk of "non-ethical terminology" to non-normative terminology, we get NNR as a claim about language. I'll understand it specifically as a claim about natural languages that can be used by human beings. A further generalization would extend it to a claim about concepts. I'll take NNR to be this more general claim.

Sturgeon's statement of NNR can be improved on in two further respects. First, the above statement makes it sound like NNR requires that we already have terminology to represent all natural properties in wholly non-normative terms. But one could well grant that we don't yet have such terminology. This is why I initially introduced NNR as the claim that nothing counts as a natural property unless it *can* be expressed, or represented, or ascribed using wholly non-normative terms or concepts. Accordingly, in what follows I'll understand NNR as saying what must be possible in principle if a property is a natural one.<sup>4</sup>

Second, Sturgeon doesn't say what counts as representing a property, or what counts as representing it in non-normative terms. Terms like 'represent', 'ascribe', and 'express' will likely function as technical terms here. Suppose for illustration that rightness is a natural property. It shouldn't be

---

<sup>3</sup> Sturgeon (2003, 536). This is a common assumption. William FitzPatrick, for instance, claims that "any tempting natural specification of the referent of 'good' will focus on something such as human needs, in which we naturally take an interest" (2008, 182). This clearly assumes NNR. FitzPatrick then notes that goodness is the sort of property that merits our interest and asks "what objective natural fact or facts would such a fact about a natural cluster property's *meriting a certain practical response* consist in?" (FitzPatrick 2008, 182). The naturalist can say 'The fact that it's good' or 'The fact that it merits such a response'.

<sup>4</sup> In some other passages, Sturgeon seems to have in mind a claim about what's possible in principle.

enough for rightness to satisfy NNR that it can be denoted by such non-normative expressions as ‘the property we’ll be thinking about in class today’ or ‘the Pope’s favorite property’.<sup>5</sup> (That normative properties can be denoted in this way doesn’t show normative non-naturalism to be false!) Giving a satisfactory account of why this should be so is tricky, though. But one intuitive difference is that the above descriptions don’t pick out rightness “in their own right” in some sense, whereas ‘right’ does.<sup>6</sup> This difference can be seen also in the following example from Matti Eklund: “Suppose that an alien linguistic community introduces into their language a word—‘thgir’—with the stipulation that ‘thgir’ is to ascribe *the property that our ‘right’ ascribes*, but this community does not in any way use their word ‘thgir’ normatively” (Eklund 2017, 75). The status of rightness as a natural property shouldn’t depend on whether a predicate like ‘thgir’ is possible. So again NNR should require that if a non-normative term or concept ‘F’ ascribes some property N, it does so in its own right. Introducing ‘thgir’ requires appeal to ‘right’, so it fails this condition.<sup>7</sup> I’m not sure just how to spell out the relevant notion of “in its own right”, but I hope the basic idea is intuitive enough. This condition on ascription or representation doesn’t imply that if NNR is true, then normative concepts or properties are reducible to ones expressible in wholly non-normative terms—at least not for any notion of reduction stronger than necessary equivalence. Nor does NNR settle by fiat the question whether the relevant notion of being normative is primarily a feature of terms and concepts, or of the facts and properties they express.<sup>8</sup>

<sup>5</sup> Jackson (1998, 119) distinguishes “denoting” a property from “ascribing” it in this kind of way.

<sup>6</sup> Whether ‘good’ or ‘right’ ascribes a normative property may vary with context. Even so, the kind of contextual input that’s involved in determining their reference looks different from that involved in determining when ‘the property we’ll be thinking about in class today’ denotes rightness.

<sup>7</sup> The same point may apply to the idea that if (as many naturalists think) normative properties play a causal role and if R is the causal role of rightness, then rightness can be represented non-normatively as ‘the property that fills causal role R’. For we may have to use ‘right’ to specify R.

<sup>8</sup> The question will pop up again in §6. For the general debate, see e.g. (Roberts 2013) and (Eklund 2017, chs. 4-5). (Finlay 2019) is a helpful overview of various

Many objections to naturalism question the possibility of ascribing normative properties in non-normative terms which satisfy these conditions. Moore's illustrations of the open question argument are like this. Another example is Derek Parfit's Triviality Objection against normative naturalism. Its core is that the central claims of normative naturalism must take the form of statements of identity between normative and natural properties, but no such claim can have all of the features, such as informativeness, required by the truth of normative naturalism (Parfit 2011, 344). One response is to argue that Parfit's objection can be met on its own terms (Dowell and Sobel 2017). But we might instead note that according to Parfit, the truth of normative naturalism requires informative identity statements of the form 'NORM=NAT', where 'NORM' is placeholder for a normative term and 'NAT' is a placeholder for a simple or complex expression that ascribes a natural property. He correctly points out that if normative naturalism is true, then 'NORM' ascribes a natural property. He also correctly points out that substituting 'NORM' for 'NAT' would make the identity statement uninformative. He concludes that an informative identity statement of this form requires that 'NAT' be a non-normative expression. So it's clear that Parfit accepts NNR, or at least attributes it to his target (see also Parfit 2011, 295). Suspend NNR, and normative naturalism doesn't require true informative identity statements of the form 'NORM=NAT'.

A more recent example is Matt Bedke's argument that normative naturalism makes normative cognition dispensable. By normative cognition, Bedke means thought and talk involving concepts or terms such as 'ought' or 'is good', whose occurrent tokenings have a special mode of presentation that involves a sense of "inherent, authoritative guidance" (Bedke 2021, 149). His worry is that normative naturalism makes this presentational quality accidental: it is "not needed to fit the job description of normative cognition—*ascribing natural properties*. That can be done with non-

---

things that 'normative' may mean when applied to concepts, judgments, properties, and more. Normative naturalists on both sides of NNR differ on whether the normative/non-normative distinction is primarily a distinction among concepts or properties.

normative (natural) cognition” (Bedke 2021, 150).<sup>9</sup> The naturalist can agree that insofar as a sense of authoritative guidance is integral to our thought and talk about how to act and live, making it dispensable would be a problem. But again, NNR is crucial for raising the problem in the first place. Suspend NNR, and non-normative cognition isn’t guaranteed to suffice for ascribing natural properties. Nor would it be an accident that if naturalism is true, normative cognition ascribes natural properties.

There are many other examples. For instance, normative naturalism is sometimes interpreted as saying that normative facts aren’t further facts relative to *non-normative* facts (Rosen 2018, 157). But all it implies is that normative facts aren’t further facts that come on the scene after the *natural* facts are fixed. This follows trivially, if normative facts are among the natural facts. I trust that readers familiar with debates over normative naturalism will recognize how widely those debates presuppose NNR. It would therefore be important to debates over normative naturalism if its truth didn’t require NNR.

### 3. Normative naturalism without non-normative representability

Normative naturalism is at its core a metaphysical thesis about the nature of normative facts and properties: they are a kind of natural facts and properties.<sup>10</sup> Naturalness in this sense isn’t a feature of words or concepts that can be used to ascribe those properties. So the core thesis of normative naturalism doesn’t involve a further thesis about the relation between two sets of terms, “normative” and “non-normative”. This is so irrespective of whether being normative and being non-normative are (primarily) features of terms or concepts, or of facts or properties. So the truth of normative

---

<sup>9</sup> This is how Bedke formulates his concern in relation to the “referential” function of normative concepts. For naturalists who appeal instead to a distinctive “non-referential” function, see the references in note 26.

<sup>10</sup> Normative naturalism might have some semantic implications. Perhaps if killing is bad and that’s a natural fact, then any sentence which represents this fact is true. This is hardly distinctive, though.

naturalism doesn't depend on NNR. Naturalists need only claim that every normative fact is *already* a natural fact, irrespective of whether there's a non-normative way of representing that fact in addition to the normative way. (In §4 I'll discuss conceptions of a natural property which allow this possibility.) Jonathan Dancy (2006, 127) dubs this view "one-term naturalism", in contrast to "two-term naturalism" which endorses the further linguistic or conceptual commitments of NNR.<sup>11</sup>

Many naturalists do adopt the two-term naturalist project of identifying which natural properties are normative properties in non-normative terms (Railton 1986; Boyd 1988; Copp 1995). It's true, but trivial, that to be wrong is to be wrong. It would be non-trivial if to be wrong were to fail to maximally promote the objective interests of everyone, impartially considered (Railton 1986), or interfere with the flourishing of societies (Copp 1995), or destabilize cooperation (Sterelny and Fraser 2017). Property identifications like these are empirical hypotheses about which natural properties the normative properties are likely to be, given certain non-normatively characterizable functions which morality and other normative codes play in human life (Isserow ms). They invite the kinds of objections canvassed above. I'm not arguing against two-term naturalism. I simply note that defending such identifications isn't necessary to the truth of naturalism. Without NNR, those objections fall away.<sup>12</sup>

Sturgeon finds NNR "highly questionable: possibly false, and at the very least requiring defense" (Sturgeon 2003, 537). He appeals to the idea that normative naturalism isn't in the first instance a doctrine about language—nor, we might add, about thought. He also notes analogies which counsel caution about NNR. One is that physicalism about the mental doesn't require that mental states be representable, even in principle, in the language

---

<sup>11</sup> Arguments for two-term naturalism might include arguments for the reducibility of normative properties (Railton 1986), arguments from supervenience (Jackson 1998), or arguments from requirements on coherent planning (Gibbard 2003). These arguments don't show that the truth of normative naturalism requires NNR.

<sup>12</sup> An individual naturalist can of course go ahead defending an informative property identification if they consider that important on some further ground. Establishing some such identification just wouldn't be necessary to the truth of normative naturalism, but more like an add-on to your basic meal deal.



of physics (Sturgeon 2003, 537). Our account of how mental states can be physical states needn't take such a form. Another concerns "metaphysical" properties, including divine goodness. If they couldn't be represented in non-normative terms, that wouldn't disqualify them from being metaphysical properties. This gives us additional reason to "ask why a property's being natural should depend on this" (Sturgeon 2003, 540). We have no reason to suppose NNR is true because of something special about representing natural properties in particular. In a different context, Sturgeon notes that scientific progress involves introducing new terms for previously unrecognized properties all the time, and it's controversial whether this process of terminological innovation has an end, even in principle (Sturgeon 2006, 99).

More generally, it seems possible for natural properties to exceed even our best representational resources, non-normative or otherwise. Any natural language can have only countably many predicates, but natural properties might not be only countably many. Some natural properties might also be more fine-grained than what natural languages or human thought can represent. Examples might include the most maximally determinate values along certain continuous physical parameters. In general, any representation abstracts from some features of its object; otherwise it duplicates rather than represents the object. At least the latter point remains even if NNR requires only that for any natural property *N*, there's a non-normative predicate in some or other natural language which ascribes *N*. These points matter because NNR seems to imply that if there are some properties which we're incapable of representing at all, those properties won't be natural. If a property's being natural depended on whether it can be represented in certain kind of way, why should properties that we can't represent be exempt from this requirement? Why not instead adopt a conception of a natural property which doesn't impose NNR-style representational conditions? We would in any case need such independent conditions for a property to be natural if we wanted to allow that the class of natural properties includes properties which we cannot represent.

The above points all concern natural properties of whatever kind, not specifically normative properties. They imply on very general grounds that whether normative properties fall into the class of natural properties doesn't

depend on how those properties relate to non-normative representations. I don't take these considerations to establish that NNR is false. The matters are complex enough that it would be folly to take them as settled. For instance, it's plausible that all of the paradigmatic natural properties that we can represent are ones we can represent in non-normative terms. Perhaps this is best explained by NNR.<sup>13</sup> People are likely to vary regarding whether the above considerations suffice to defeat that inference. But for my purposes I don't need to show that NNR is false, but only that it's questionable enough for normative naturalists to have reason to keep their distance since they can. I take the above to show this much. There may also be further worries about NNR as applied specifically to normative properties. One example would be if naturalists thought that some normative properties, such as wrongness, are somehow essentially normative. (However, such views are more commonly raised as *objections* to normative naturalism. I'll return to this in §5.) Another potential example is the view that the extensions of normative terms and concepts aren't unified under non-normative similarity relations. If this "shapelessness thesis" is compatible with normative naturalism, that might be another reason to worry about NNR.<sup>14</sup>

#### 4. Naturalness and non-normative representability

One concern about divorcing normative naturalism from NNR is that we might lose our grip on the sense in which normative properties are supposed to be natural. How are normative properties supposed to fall into the class of natural properties, if not in virtue of how normative properties are related to properties which are fairly uncontroversially natural and can be ascribed in non-normative terms? The main issue for my purposes is this:

---

<sup>13</sup> Thanks to a reviewer for this journal for pressing this response. They also worried that the most salient naturalistic accounts of reference-determination support a case for NNR. I don't think the issue is nearly as clear-cut, but cannot address this properly for reasons of space. (For one relevant point, see the end of §6.)

<sup>14</sup> Väyrynen (2014) argues that the shapelessness thesis is compatible with ethical naturalism, doesn't require normative particularism, and can be explained by more general factors not specific to the normative.

one-term naturalism requires a conception of a natural property on which normative properties meet the conditions for naturalness directly, rather than in virtue of how our normative ways of representing them are related to non-normative representations. The good news is that none of the three most prominent accounts in metaethics of which properties are natural properties implies that any plausible normative naturalism must accept NNR.<sup>15</sup>

Suppose that any property that is such as to play a causal role in the natural world (or else figure in causal explanations of events or states of affairs) is a natural property. Whether something plays a causal role doesn't depend on how we describe it. So normative properties like goodness and wrongness qualify as natural in their own right if they play a causal role in the world, even if they cannot thereby be represented in wholly non-normative terms. Sturgeon suggests that "placing a property in a causal network is a way of saying something about which property it is, even if one lacks an explicit reduction for it" (Sturgeon 2006, 100). It's of course controversial whether normative properties meet this condition or its stronger sibling which requires an *ineliminable* causal or explanatory role.<sup>16</sup> It's also controversial whether the satisfaction of the relevant explanatory condition by normative properties entails the metaphysical claims of normative naturalism (Sinclair 2011). But these aren't debates about whether a plausible normative naturalism must accept NNR.

Or, suppose a natural property is such that synthetic propositions about its instantiation aren't strongly *a priori* but are subject to empirical constraint (Copp 2003, 181; Boyd 1988). If normative properties met this condition, they would do so in their own right: the propositions to be tested would be propositions involving normative concepts. Specifying how propositions about the instantiation of a property can come to be known, and how their justification may be defeated, is again a way of saying something about which property it is. Whether our basic moral knowledge is strongly *a priori* is a familiar debate, of course. But it's not a debate about whether a plausible normative naturalism must accept NNR.

---

<sup>15</sup> Copp (2003) and Väyrynen (2009) survey the main options relevant to meta-ethical debates.

<sup>16</sup> For a classic exchange, see Sturgeon (1985; 1986) vs. Harman (1986).

Or, suppose that any property posited in the best scientific accounts of the world is natural (Shafer-Landau 2003, 59). In the case of morality, naturalists argue that there is an empirical discipline which deals with ethical matters and is no less apt to figure in the best scientific accounts of the world than psychology or sociology—namely, a discipline called ethics (Boyd 1988, 206-8; Sturgeon 2003, 553).<sup>17</sup> If ethics had such a disciplinary status, then establishing principles linking non-normative properties to normative ones through first-order normative inquiry would be a way of saying something about which natural properties goodness and rightness are even if those connections aren't so robust as to satisfy NNR. It's of course controversial whether ethics has this kind of disciplinary status. But that's again not a debate about whether a plausible normative naturalism must accept NNR.

I conclude that the most prominent accounts of natural properties in debates about normative naturalism and non-naturalism don't imply that we'll lose our grip on the notion of a natural property if plausible forms of normative naturalism needn't accept NNR. So no new reason has emerged to treat NNR as a condition on the truth of normative naturalism.

## 5. The “just too different” objection

I'll now turn to the “just too different” (JTD) objection to normative naturalism. The objection has it that the things we represent in normative ways are, intuitively, just too different from the things we represent in non-normative ways for them to be metaphysically of a kind. As Dancy puts it: “There remains a stubborn feeling that [normative] facts about what is right or wrong, what is good or bad, and what we have reason to do have something distinctive in common, and that this common feature is something

---

<sup>17</sup> Several normative naturalists suggest that *health* is an evaluative concept which picks out a property that plays genuine explanatory roles (Bloomfield 2001; Sturgeon 2003, 553; Railton 2018, 51) and so there are uncontroversially naturalistic disciplines that deal with questions of value. This may require a view of “thick” concepts which is widely endorsed but which I myself find questionable (Väyrynen 2013; cf. Cline 2015).

that a natural fact could not have” (Dancy 2006, 136).<sup>18</sup> For instance, nothing can count as good or right unless it’s something that we *ought* to be concerned to promote, or *merits* being given a certain kind of weight in deliberation, or the like. Paradigmatic natural facts—ranging from facts of physics and chemistry to non-normative facts about the colors of objects, the needs and desires of human and non-human animals, and the like—aren’t like that. So why think that facts with the kind of special importance that normative facts seem to have are metaphysically of a kind with paradigmatic non-normative natural facts which lack such importance?

The JTD objection can be raised against both one-term naturalism and two-term naturalism, since it relies on a contrast between properties represented in normative terms and properties represented in paradigmatically non-normative terms. But I suspect the objection derives some of its force from assuming that normative naturalism is committed to NNR. An intuitive contrast between paradigmatic non-normative natural properties and properties like rightness and wrongness is less compelling as an objection to normative naturalism if properties can be natural without conforming to non-normative paradigms like fermentation, color, need, or desire. That paradigmatic members of class C lack feature F does little by itself to show that C doesn’t have a subclass whose members do have F. The residual force of the objection depends on what counts as the kind of normative importance which is supposed to set normative facts apart from (other) natural facts. The more distinctive normative facts are from these other facts in this respect, the more forceful the concern.

Words like ‘good’, ‘right’, and ‘ought’ are often used to express forms of normativity which are naturalistically acceptable. These include norm-relative normativity characteristic of conventional norms (law, etiquette) and role obligations (such as what’s required of teachers), kind-relative normativity (such as being a good toaster), and instrumental normativity.<sup>19</sup>

---

<sup>18</sup> See also Nagel (1986, 138), FitzPatrick (2008, 179-82), Enoch (2011, 104-8), and Parfit (2011, 324-27). I won’t be able to do justice to various nuances that can be found in these and other discussions of the just-too-different objection. For a helpful survey of the debate, see Paakkunainen (2017).

<sup>19</sup> See e.g. (Paakkunainen 2017, 3) and the references therein. Not everyone thinks that these forms of non-categorical normativity are less puzzling than categorical

These all involve standards such that if you fail to satisfy them, you're open to a certain kind of criticism. So if the special features that are supposed to make it implausible that normative properties fall into the class of natural properties were exhibited also by these normative properties, naturalists needn't worry. It's thus no surprise that the JTD objection tends to focus on a subclass of normative notions, such as being morally right or wrong, what one has normative reason to do, what one really ought to do, and the like. Their normativity has struck many as more "genuine" or "authoritative" than these other kinds of normativity.

Genuine or authoritative normativity is often characterized in terms of a distinctive role. Authoritatively normative concepts or judgments play some characteristic or essential role in deliberation. Authoritatively normative facts and properties have some characteristic or essential connection to decision and action. Proposals vary in terms of whether such connections are themselves normative. Either way, they are supposed to be different in kind from how both non-authoritative forms of normativity and non-normative notions may relate to deliberation and action. But it has proved difficult to pin down just what connections are meant to characterize robust or authoritative normativity. As Hille Paakkunainen notes: "There's currently little agreement on hallmarks of genuine normative importance—beyond, perhaps, certain intranormative connections between important normative notions" (2017, 9).

A common way to illustrate the JTD objection is to say that some normative facts, such as moral facts, are intrinsically significant in that any rational agent will have normative reason to respond accordingly. If such categorical reasons were a hallmark of the normative authority of morality, then normative naturalism would seem hard pressed to account for its authority. For many think that this kind of categorical normativity isn't compatible with a naturalistic world view. But we should distinguish two questions here. One is whether it's true that, no matter what the moral facts are like and what moral agents and their environments are like, any moral agent will by necessity have normative reason to act morally. The other is

---

normativity. How exactly this might relate to normative naturalism is a more complex issue than I can address here, however.

whether the idea that moral facts are categorically reason-giving is a firm datum whose denial automatically implies a significant loss of plausibility.

Even if it's true that moral facts in fact are necessarily reason-giving, that claim amounts to a substantive theoretical position, not a pre-theoretical datum. The claim that moral facts are reason-giving is logically weaker than the claim that they are necessarily so. Why then think the latter is a firm default? Many naturalists argue that genuinely pre-theoretical data about the importance that our normative practices assign to moral facts can be accounted for even if our reasons to be moral obtain contingently. Such explanations typically take the following form. Given (i) some plausible assumptions about what kind of social and emotional factors are robust features of human social environments and psychology and (ii) some plausible first-order moral assumptions, it's a robust empirical generalization that moral agents have normative reason to do what's good and avoid what's bad (Brink 1984; Railton 1986; Boyd 1988; Copp 1995; Isserow ms). The kind of reasons internalism that underpins these explanations doesn't imply that our reasons for doing what morality tells us to do are merely instrumental (cf. Williams 1981). The contingency of such reasons also needn't be contingency on fragile preferences or desires. The relevant generalization may break down in anomalous cases—ideally coherent Caligulas or the like. But such individuals would be so far removed from most of us that it's unclear why their existence should be a threat to the normative authority of morality. It wouldn't be accidental that most of us, most of the time, have reason to do what morality tells us to do.

My aim here is to indicate how naturalists explain the importance that our normative practices assign to moral facts in terms of reasons to be moral that obtain as a matter of robust and deep contingency, not to defend these accounts. What I want to highlight is that such accounts can satisfy certain independently plausible conditions on normative authority which require less than treating moral facts as necessarily reason-giving. For instance, William FitzPatrick suggests that the significance of what various forms of non-normative inquiry (such as biology, psychology, and sociology) can contribute to moral inquiry “must be assessed through the lens of autonomous ethical reflection on our life and experience” because “nothing presented to a rational agent in any other way could be *authoritative* for her”

(FitzPatrick 2008, 172). Normative naturalists can well accept that assessing the significance of potential inputs into moral inquiry is a central task of normative thought and talk. Especially if we don't rely on NNR, non-normative representations of natural properties may not suffice for this task. Any field of inquiry seems "theory-dependent" in its reliance on auxiliary theoretical assumptions, including some drawn from that very field (Boyd 1988, 190, 207). For a simple example in ethics, consider a modest moral principle: no morally admirable person would instigate and oversee the deaths of millions of people. And consider a moral claim which some people accept: Hitler was a morally admirable person. These two moral claims jointly yield an empirical consequence: Hitler didn't instigate and oversee the deaths of millions of people. But it's an empirical fact that he did. So we know, on the basis of empirical test, that at least one of these moral claims must be rejected.<sup>20</sup> The empirical constraint doesn't say which one to reject. But that's par for the course: claims are assessed in bundles, not singly in isolation. So naturalists can agree that autonomous ethical reflection is required to assess which moral claims we should reject.

A deadlock now threatens the debate about the JTD objection. It's dialectically inadmissible for the objection to presuppose that moral facts are necessarily or categorically reason-giving. But the objection fails under various weaker notions of authoritative normativity, such as those characterized in terms of deliverances of autonomous normative reflection. Naturalists may also be able to accommodate a robust sense in which genuine norms are inescapable. (For instance, they might adapt proposals from Woods 2018.) And they can accept that wrongness, for instance, is authoritatively normative in the sense that it involves violations of important standards which warrant blame, other things being equal (Copp 2020a). Whether normative properties are just too different from natural properties depends on what's packed into a notion of authoritative normativity which it would be dialectically admissible for the JTD objection to deploy. The jury's still out on that.

---

<sup>20</sup> Nick Sturgeon used this example in an undergraduate lecture on normative ethics which I once sat in on.



## 6. Authoritative normativity in thought and talk

I'll now offer a slightly sideways approach to the JTD objection. I'll sketch a kind of proof of concept that normative naturalism faces no distinctive trouble making sense of thought and talk involving authoritative normativity. I'll do this by showing how naturalists can make sense of a certain type of authoritatively normative concept if they so desire. Whatever trouble naturalism may face will be of a sort faced by other metaethical views as well. While I hope that the concept I'll focus on is representative, I won't be able to show that the strategy generalizes to further notions of authoritative normativity that I haven't discussed.<sup>21</sup>

This strategy speaks to the JTD objection even though the objection is normally framed in terms of normative facts and properties rather than terms or concepts. As noted in §2, there's a dispute about whether being normative is primarily a feature of terms and concepts, or of the facts and properties they express. On the one hand, if normative properties are normative because of some features of the terms or concepts that ascribe them, then making naturalist sense of authoritative normativity is primarily a task of making naturalist sense of thought and talk involving authoritative normativity. If normative properties instead are normative because of some features of non-representational reality, we would still expect our thought and talk about such properties somehow to reflect whatever connection to decision and action marks a fact or property as authoritatively normative. Either way, if naturalists can make sense of thought and talk involving authoritative normativity, then nothing in its nature shows that the properties ascribed by such thought and talk are just too different from (other)

---

<sup>21</sup> Another limitation is that I set aside first-order questions about which standards or facts are authoritatively normative. For one such first-order account, see (Rowland forthcoming). That account seems compatible with normative naturalism. A broader question here is whether it's better to think of authoritatively normative oughts (if there are any) as exemplifying a distinct ought-concept (or concepts) or as combining an exemplification of some independently possessable ought-concept (MORAL OUGHT, or the like) with a higher-order property of being authoritatively normative in some sense (for this distinction, see (Howard and Laskowski ms)). I hope my discussion to be modifiable to fit either model.

natural properties to be natural properties themselves. This should also enable naturalists to explain why such thought and talk might be not only indispensable for thinking about authoritatively normative facts, but also important. The strategy is available to both one-term and two-term naturalists. (The latter would need to show that normative concepts are indispensable for, for example, deliberation or normative knowledge even though everything that can be said in normative terms about what the world is like can also in principle be said in wholly non-normative terms.<sup>22</sup>) But it comes with one less headache to normative naturalists who aren't committed to NNR: making naturalist sense of thought and talk involving authoritative normativity doesn't require showing that the properties it ascribes can in principle be also ascribed in wholly non-normative terms.

I'll explain the strategy I have in mind in terms of an authoritatively normative concept individuated by its distinctive role in the deliberative activity of non-arbitrary selection. This is the concept Tristram McPherson labels PRACTICAL OUGHT: a normative concept which has distinctive authority because it's "the concept of a norm which is *the norm to appeal to* in the context of non-arbitrary selection" (McPherson 2018, 267; cf. McPherson 2020).<sup>23</sup> When morality requires one thing but prudence requires something else, a resolution of their relative importance had better be non-arbitrary, and a norm that provides such a solution would seem to lord over the norms whose conflict it resolves. This might be the concept that some call the concept of ought *simpliciter*.<sup>24</sup> It looks like a dialectically admissible tool for assessing whether normative naturalism can make sense of how authoritative normativity figures in our normative thought and talk.

Normative naturalism has no distinctive trouble making sense of how this sort of authoritative normativity figures in language. The talk you

---

<sup>22</sup> This is the standard two-term naturalist strategy. Its most common form is to say that normative concepts are distinguished by a kind of conceptual role which non-normative concepts are of a wrong kind to serve.

<sup>23</sup> I use small caps to denote concepts. To be clear, my discussion here concerns the concept McPherson (2018) dubs PRACTICAL OUGHT, not his particular analysis of it.

<sup>24</sup> Some philosophers are skeptical of ought *simpliciter* (Baker 2018; Copp 2020b; Howard and Laskowski ms). If they're right, normative naturalism faces no challenge from authoritative normativity in this sense.

sometimes see of “the special nature of normative words” is misleading. There are no normative words (not in English anyway) in the strong sense of words that are conventionally associated with, specifically, authoritative normativity. Rather, there are words that can be used normatively, in various senses of ‘normative’. Which sort of claim a given assertive utterance of a word like ‘ought’, ‘wrong’, or ‘good’ expresses depends on the context of utterance, in a potentially complex sort of way.<sup>25</sup> In some contexts ‘ought’ expresses PRACTICAL OUGHT. But in many contexts it expresses forms of non-authoritative normativity that are (as we saw above) widely agreed to be naturalistically acceptable. The latter entails that the context-invariant features of what ‘ought’ means are compatible with normative naturalism. Nor does normative naturalism conflict with features specific to contexts in which ‘ought’ expresses PRACTICAL OUGHT, such as the notion of “the norm to appeal to” or the notion of non-arbitrary selection.

Getting the relevant ought into thought is slightly more fraught. One option is that authoritative normativity figures in ways of thinking about normative properties which don’t amount to distinct concepts or modes of presentation for them but instead play some indispensable non-referential function.<sup>26</sup> I’ll instead explain how naturalists have no distinctive trouble making sense of a distinct concept like PRACTICAL OUGHT. Here’s the key point: any concept like PRACTICAL OUGHT must involve an ordering of the items that are relevantly live options in the given context. The ought-structure is in general such that what you ought to do is one among the top options on the contextually relevant ordering of the relevantly live options. That’s what falls out of the standard sort of descriptive semantics for ‘ought’ in deontic contexts, broadly in the vein of Kratzer (1991). There’s no reason to think that PRACTICAL OUGHT is an exception. ‘Ought’ can be used to express it. More importantly, if PRACTICAL OUGHT didn’t involve an ordering, it couldn’t play its role in non-arbitrary selection when requirements of

---

<sup>25</sup> In not saying more about this, I’m skipping many complex issues regarding the semantics and metasemantics of these terms and their context-sensitivity. In other work I argue that the practical role of words like ‘ought’ isn’t a feature of their descriptive semantics or metasemantics (Väyrynen forthcoming).

<sup>26</sup> For such accounts, see (Copp 2018; 2020a) and (Laskowski 2019). For some objections, see (Bedke 2021).

morality conflict with the law, when promoting the interests of one's beloved conflict with fairness, and so on. So a non-defective application of PRACTICAL OUGHT commits the thinker to there being a certain kind of ordering which somehow gets uniquely selected by the concept or its application in a context. (Uniqueness is required if that ordering is to be *the* norm to appeal to in non-arbitrary selection. In what follows, I'll simplify presentation by assuming that uniqueness is baked into non-arbitrariness.)

What ordering is this? Ought-concepts are individuated in part (though perhaps not wholly) by the orderings they involve. MORAL OUGHT and PRUDENTIAL OUGHT are different concepts if (though perhaps not only if) they rank options by different criteria. Structurally speaking PRACTICAL OUGHT looks no different; it's individuated in part in relation to an ordering (whichever it is) that provides a basis for successful non-arbitrary selection among the relevantly live options in the context. Content-wise, PRACTICAL OUGHT doesn't encode a specific ordering or its source. But MORAL OUGHT doesn't do so either; it's a substantive question what the correct moral standards are. So it's hardly distinctive of PRACTICAL OUGHT that specifying how the relevantly live options rank will require substantive normative inquiry and is subject to dispute and disagreement. Ought-structure in general and PRACTICAL OUGHT in particular may also do little to restrict what considerations may coherently be treated as relevant to what one practically ought to do. This concept's contribution to the content of thoughts that token it may accordingly be informatively fairly thin—something like a condition characterizable as *ranking highly on an ordering, whichever it is, which provides a basis for successful non-arbitrary selection among the relevantly live options*.

These are wholly general points about PRACTICAL OUGHT. As broadly structural points, they're largely neutral between a wide range of metaethical views, realist and antirealist alike. This should already lead us to expect that normative naturalism should have no distinctive trouble accounting for PRACTICAL OUGHT thoughts. But to check this, let's look at issues where different metaethical views might differ, ontology aside. There are issues in the philosophy of mind, such as what kind of mental state someone's in when they judge that they practically ought to do something. And there are issues in the metasemantics of normative concepts, such as in virtue of

what factors PRACTICAL OUGHT comes to pick out the property of ranking highly among a set of live options on the relevant kind of ordering.<sup>27</sup>

As regards philosophy of mind, most naturalists would characterize the judgment that I practically ought to  $\varphi$  as a belief that  $\varphi$ -ing ranks highly on the relevant kind of ordering. It isn't clear why there should be any deep puzzle as to how such a belief could serve the deliberative role of PRACTICAL OUGHT.<sup>28</sup> If I believe that I practically ought to do something, I'm committed to thinking that a certain resolution to my practical situation is correct. That's just built into the ordering which induces that resolution. It's a further question whether I'm genuinely committed let alone motivated to act that way. Just as sharing a normative concept may not require a lot of uniformity in inputs to its application, it may not require a lot of uniformity in the practical upshots of its application (cf. Merli 2009).

As regards metasemantics, normative naturalists of course have work to do in explaining the reference of concepts like PRACTICAL OUGHT. If PRACTICAL OUGHT stands for ranking highly on a certain kind of ordering, there are questions about whether that property is natural and how PRACTICAL OUGHT comes to pick it out. Is its reference fixed by how its use is causally regulated, or its conceptual role, or the functions of the representational systems that use the concept, or some combination of these or other factors? How does the relevant mechanism fit with what makes a property natural, whether that be playing a causal role, being empirical, or something else? (This matters especially to one-term naturalists who need normative properties to meet the conditions for naturalness in their own right.) But these questions aren't special to PRACTICAL OUGHT. Whether and how normative concepts get to pick out natural properties just is the general metasemantic question for naturalism, just as whether and how they get to pick out non-natural properties is the general metasemantic question for non-naturalism, and likewise for other views in normative metaphysics.

It's not clear why normative naturalists should be in any worse position than others in accounting specifically for PRACTICAL OUGHT thoughts. One

---

<sup>27</sup> Or, perhaps, a property meeting a condition so characterized. I won't distinguish these below for simplicity.

<sup>28</sup> Thus, I don't see why McPherson (2020, 1344-45) thinks there's a deep puzzle here for normative realists.

consideration here is that this concept doesn't commit us to categorical reasons. The claim that successful non-arbitrary selection necessarily provides normative reasons for action is a substantive claim not built into the content of PRACTICAL OUGHT. Another consideration is that normative naturalism looks no worse off in ensuring that PRACTICAL OUGHT has an acceptably determinate representational content.<sup>29</sup> For this concept to have a non-empty extension on *any* view, there must be an ordering which provides a basis for successful non-arbitrary selection among the relevantly live options in the context. Some vagueness aside, this should ensure sufficient determinacy in representational content. At minimum, it is not clear why failure of uniqueness should turn on whether this relation is a natural relation. That the content of the ordering is subject to ignorance, uncertainty, and disagreement also doesn't mean that PRACTICAL OUGHT thoughts lack acceptably determinate content. Finally, if PRACTICAL OUGHT stands for ranking highly on a certain kind of ordering, normative naturalists needn't worry about whether it can be represented in wholly non-normative terms unless they accept NNR. They can also agree that determining which properties normative concepts refer to may require first-order normative assumptions.<sup>30</sup> There may be no normatively neutral way to determine whether GOOD refers to *goodness* even if goodness is a natural property, and likewise for PRACTICAL OUGHT.

In saying that it's not clear why normative naturalists should be any worse off here, I really don't mean to be bloody-minded. It remains a live question whether there is a natural property or relation which satisfies a condition like *ranks highly on an ordering, whichever it is, which provides a basis for successful non-arbitrary selection among the relevantly live options in the given deliberative context*, and does so in such a way that

---

<sup>29</sup> McPherson (2020, 1346-51) argues that ensuring representational determinacy is a serious challenge to normative realists. Here I can pick up on only one dimension of the challenge he poses. However, some of the dimensions I bracket strike me as less troublesome, since McPherson takes as his foil an overly simple sort of causal metasemantics (simpler, say, than the epistemically constrained account in (Boyd 1988)). But see (Schroeter and Schroeter 2013).

<sup>30</sup> I discuss this briefly in relation to Boyd's causal metasemantics in (Väyrynen 2019, 206-8).

PRACTICAL OUGHT can be said to stand for that property. My point here is that there's no good reason why this sort of authoritatively normative property should be just too different from (other) natural properties to itself be natural, given that normative naturalism has no distinctive trouble making sense of thoughts involving it. It remains to be seen whether this strategy generalizes to other relevant authoritatively normative concepts besides PRACTICAL OUGHT. But on this proof of concept, issues facing normative naturalism wouldn't be distinctive. They would be just the same general issues that face any metasemantics for normative thought and talk.

## 7. Conclusion

My aim in this paper has been to contribute to debates about how normative properties might figure in our thought and talk if normative naturalism is true. First, I offered some improvements on Sturgeon's formulation of an assumption about representation of natural properties which I called Non-Normative Representability, and noted several objections to normative naturalism which presuppose NNR. Second, I offered some reasons, Sturgeon's and my own, why the truth of normative naturalism doesn't require NNR and why NNR is questionable enough for naturalists to have reason to keep their distance since they can. If that's right, the objections in question fall away. Third, I offered a slightly sideways approach to the "just too different" objection to normative naturalism. I first suggested that the objection loses some of its force if normative naturalism isn't committed to NNR and noted difficulties in specifying the notion of "genuine" or "authoritative" normativity which is the objection's primary concern. I then tried to make progress with a proof of concept that naturalists have no distinctive trouble making sense of thought and talk involving the relevant kind of authoritative normativity. While the strategy is compatible with NNR, its execution will prompt fewer headaches if normative naturalists don't count NNR among their commitments. I leave it for future work to assess how well the strategy I offer generalizes.<sup>31</sup>

---

<sup>31</sup> I dedicate this paper to the memory of Nick Sturgeon, who was a member of my dissertation committee at Cornell University. I wrote it to highlight the significance

---

## References

- Baker, Derek. 2018. "Skepticism About Ought *Simpliciter*." *Oxford Studies in Metaethics* 13: 230–52. DOI:10.1093/oso/9780198823841.003.0011
- Bedke, Matthew S. 2021. "Naturalism and Normative Cognition." *Philosophical Studies* 178: 147–67. <https://doi.org/10.1007/s11098-020-01425-y>
- Bloomfield, Paul. 2001. *Moral Reality*. Oxford University Press.
- Boyd, Richard N. 1988. "How to Be a Moral Realist." In *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord, 181–228. Cornell University Press.
- Brink, David O. 1984. "Moral Realism and the Sceptical Arguments from Disagreement and Queerness." *Australasian Journal of Philosophy* 62 (2): 111–25. <https://doi.org/10.1080/00048408412341311>
- Cline, Brendan. 2015. "Moral Explanations, Thick and Thin." *Journal of Ethics and Social Philosophy* 9 (2): 1–21. <https://doi.org/10.26556/jesp.v9i2.89>
- Copp, David. 1995. *Morality, Normativity, and Society*. Oxford University Press.
- Copp, David. 2003. "Why Naturalism?" *Ethical Theory and Moral Practice* 6: 179–200. <https://doi.org/10.1023/A:1024420725408>
- Copp, David. 2018. "Realist Expressivism and the Fundamental Role of Normative Belief." *Philosophical Studies* 175: 1333–56. <https://doi.org/10.1007/s11098-017-0913-6>
- Copp, David. 2020a. "Just Too Different: Normative Properties and Natural Properties." *Philosophical Studies* 177: 263–86. <https://doi.org/10.1007/s11098-018-1189-1>
- Copp, David. 2020b. "Normative Pluralism and Skepticism About 'ought *Simpliciter*'." In *The Routledge Handbook of Practical Reason*, edited by Ruth Chang and Kurt Sylvan, 416–37. Routledge.
- Dancy, Jonathan. 2006. "Nonnaturalism." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 122–45. Oxford University Press.
- Dowell, J. L., and David Sobel. 2017. "Advice for Non-Analytic Naturalists." In *Reading Parfit: On What Matters*, edited by Simon Kirchin, 153–71. Routledge.
- Eklund, Matti. 2017. *Choosing Normative Concepts*. Oxford: Oxford University Press.
- Enoch, David. 2011. *Taking Morality Seriously*. Oxford: Oxford University Press.

---

of some of Nick's observations about the commitments of normative naturalism which its critics often neglect. I'm grateful to David Copp, Camil Golub, Jessica Isserow, Gerald Lang, Richard Rowland, a work-in-progress group at University of Leeds, the participants of the *Value in Language* workshop, and two anonymous reviewers for helpful comments on earlier drafts, and to Dan Zeman for inviting me to contribute.



- Finlay, Stephen. 2019. "Defining Normativity." *Dimensions of Normativity. New Essays on Metaethics and Jurisprudence*, edited by David Plunkett, Scott J. Shapiro, and Kevin Toh, 187–219. Oxford: Oxford University Press.
- FitzPatrick, William, J. 2008. "Robust Ethical Realism, Non-naturalism, and Normativity." *Oxford Studies in Metaethics* 3: 159–205.
- Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Harman, Gilbert. 1986. "Moral Explanations of Natural Facts: Can Moral Claims be Tested Against Moral Reality?" *The Southern Journal of Philosophy* 24 (S1: *Spindel Supplement: Moral realism*, edited by Norman Gillespie): 57–68.  
<https://doi.org/10.1111/j.2041-6962.1986.tb01596.x>
- Howard, Nathan, Nicholas Laskowski. ms. "No Gods, No Masters, No Authoritative Normativity." Unpublished manuscript.
- Isserow, Jessica. ms. "Naturalising Moral Naturalism." Unpublished manuscript.
- Jackson, Frank. 1998. *From Metaphysics to Ethics*. Oxford University Press.
- Kratzer, Angelika. 1991. "Modality." In *Semantics*, edited by Arnim von Stechow, and Dieter Wunderlich, 639–50. De Gruyter.
- Laskowski, Nicholas. 2019. "The Sense of Incredibility in Ethics." *Philosophical Studies* 176: 93–115. <https://doi.org/10.1007/s11098-017-1007-1>
- McPherson, Tristram. 2018. "Authoritatively Normative Concepts." *Oxford Studies in Metaethics* 13: 253–77. DOI:10.1093/oso/9780198823841.003.0012
- McPherson, Tristram. 2020. "Deliberative Authority and Representational Determinacy: A Challenge for the Normative Realist." *Ergo* 6 (45): 1331–58.  
<https://doi.org/10.3998/ergo.12405314.0006.045>
- Merli, David. 2009. "Possessing Moral Concepts." *Philosophia* 37: 535–56.  
<https://doi.org/10.1007/s11406-009-9180-x>
- Moore, G. E. 1903. *Principia Ethica*. Cambridge University Press.
- Nagel, Thomas. 1986. *The View from Nowhere*. Oxford University Press.
- Paakkunainen, Hille. 2017. "The "Just Too Different" Objection to Normative Naturalism." *Philosophy Compass* 13 (2): 1-13. <https://doi.org/10.1111/phc3.12473>
- Parfit, Derek. 2011. *On What Matters* (Vol. 2). Oxford University Press.
- Railton, Peter. 1986. "Moral Realism." *The Philosophical Review* 95 (2): 163–207.  
<https://doi.org/10.2307/2185589>
- Railton, Peter. 2018. "Naturalistic Realism in Metaethics." In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson, and David Plunkett, 43–57. Routledge.
- Roberts, Debbie. 2013. "It's Evaluation, Only Thicker." In *Thick Concepts*, edited by Simon Kirchin, 78-96. Oxford University Press.
- Rosen, Gideon. 2018. "Metaphysical Relations in Metaethics." In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson, and David Plunkett, 151–69. Routledge.

- Rowland, Richard. forthcoming. "The Authoritative Normativity of Fitting Attitudes." *Oxford Studies in Metaethics* 17.
- Shafer-Landau, Russ. 2003. *Moral Realism*. Oxford University Press.
- Schroeter, Laura and Schroeter, François. 2013. "Normative Realism: Co-reference without Convergence?." *Philosophers' Imprint* (13): 1–24. <http://hdl.handle.net/2027/spo.3521354.0013.013>
- Sinclair, Neil. 2011. "The Explanationist Argument for Moral Realism." *Canadian Journal of Philosophy* 41 (1): 1–24. <https://doi.org/10.1353/cjp.2011.0005>
- Sterelny, Kim, and Ben Fraser. 2017. "Evolution and Moral Realism." *The British Journal for the Philosophy of Science* 68 (4): 981–06. <https://doi.org/10.1093/bjps/axv060>
- Sturgeon, Nicholas L. 1985. "Moral Explanations." In *Morality, Reason and Truth. New Essays on the Foundations of Ethics*, edited by David Copp, and David Zimmerman, 49–78. Rowman & Allanheld.
- Sturgeon, Nicholas L. 1986. "Harman on Moral Explanations of Natural Facts." *The Southern Journal of Philosophy* 24 (S1: *Spindel Supplement: Moral Realism*, edited by Norman Gillespie): 69–78. <https://doi.org/10.1111/j.2041-6962.1986.tb01597.x>
- Sturgeon, Nicholas L. 2003. "Moore on Ethical Naturalism." *Ethics* 113 (3): 528–56. <https://doi.org/10.1086/345627>
- Sturgeon, Nicholas L. 2006. "Ethical Naturalism." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 91–121. Oxford University Press.
- Väyrynen, Pekka. 2009. "Normative Appeals to the Natural." *Philosophy and Phenomenological Research* 79 (2): 279–314. <https://doi.org/10.1111/j.1933-1592.2009.00279.x>
- Väyrynen, Pekka. 2013. *The Lewd, the Rude and the Nasty. A Study of Thick Concepts in Ethics*. Oxford University Press.
- Väyrynen, Pekka. 2014. "Shapelessness in Context." *Noûs* 48 (3): 573–93. <https://doi.org/10.1111/j.1468-0068.2012.00877.x>
- Väyrynen, Pekka. 2019. "Normative Commitments in Metanormative Theory." In *Methodology and Moral Philosophy*, edited by Jussi Suikkanen, and Antti Kauppinen, 193–213. Routledge.
- Väyrynen, Pekka. forthcoming. "Practical Commitment in Normative Discourse." *Journal of Ethics and Social Philosophy*.
- Williams, Bernard. 1981. "Internal and External Reasons." In *Moral Luck*, 101–13. Cambridge University Press.
- Woods, Jack. 2018. "The Authority of Formality." *Oxford Studies in Metaethics* 13: 207–29. DOI:10.1093/oso/9780198823841.003.0010

## Illocutionary Disagreement in Faultless Disagreement

Natalia Karczewska\*

Received: 3 December 2020 / 1<sup>st</sup> Revised: 22 May 2021 /  
2<sup>nd</sup> Revised: 19 July 2021 / Accepted: 8 August 2021


*Abstract:* The debates over the problem of faultless disagreement have played a major role in shaping the landscape of today's semantic theories. In my paper, I argue that even though the existent contextualism-friendly proposals explain a lot of disagreement data by specifying various ways in which speakers may use subjective predicates, neither provides a satisfactory account which would explain what all the subjective disagreements have in common. In particular, what is lacking is an explanation of the persistent autocentric cases (Lassersohn 2004), i.e., disagreements in which each speaker utters a subjective sentence while openly and knowingly occupying his or her own perspective. In my paper, I offer a solution which consists in supplementing the standard contextualist semantics with an explanation of this most problematic class of cases, which is possible due to re-describing the phenomena in speech act nomenclature.

*Keywords:* Contextualism; commitment; faultless disagreement; speech act theory; value terms.

---

\* University of Warsaw

 <https://orcid.org/0000-0003-0889-7169>

 Krakowskie Przedmieście 26/28, 00-927 Warszawa, Poland.

 [natalia.karczewska@uw.edu.pl](mailto:natalia.karczewska@uw.edu.pl)



## 1. Introduction

Faultless disagreement (FD), as the name itself suggests, is such a conversation that gives rise to two intuitions: (1) that the speakers are disagreeing and (2) that neither of the speakers has made a mistake in uttering what they have uttered (Kölbel 2004, Lasersohn 2005). Such impressions arise, among other situations,<sup>1</sup> in relation to discussions about aesthetic, moral or gustatory value, e.g.:

[Dialogue 1]

Amy: Brussels sprouts are tasty.

Betty: No, Brussels sprouts are not tasty at all.

These two intuitions are arguably in a kind of tension—generally, on some pre-theoretical construal of the terms ‘fault’ and ‘disagreement’, if two speakers really disagree (that is, they are not mistaken about what the other one is saying and the dispute is not merely verbal), one of them must be wrong. If nobody is wrong, then the dispute cannot be a real disagreement (Boghossian 2006, Glanzberg 2007, Stojanovic 2007). This tension is particularly noticeable in discussions concerning the objective realm, where disagreement is typically construed as a situation in which one speaker expresses some proposition *p*, while the other expresses either the negation of this very proposition or some proposition which entails its negation. Faultlessness, on the other hand, is understood as saying something true—that is, something that is in agreement with the facts. Clearly then, in discussions about *objective* facts, faultlessness entails lack of disagreement and *vice versa*, so in any given situation either the speakers are disagreeing or they are faultless, but not both. Many authors believe, however, that value disagreements are different in this respect. In their case no priority can be given to either of the competing intuitions and, consequently, an adequate account of evaluative language should be able to explain both of them. The problem is that in many cases, a theory which provides a plausible account of the semantic content of value sentences, which, in my opinion, requires

---

<sup>1</sup> Some authors believe that FD intuitions arise also in conversations involving vague descriptive predicates or epistemic modals (e.g., Dietz 2008, Richard 2008, Stephenson 2007).

also the ability to account for the faultlessness appearances, has trouble explaining the disagreement intuitions in equally robust semantic terms. Since (1) and (2) are difficult to reconcile within standard accounts of the semantics of such sentences, faultless disagreements have constituted a much-discussed challenge to contemporary theories of meaning.

It seems that nowadays there are three main contenders that, sometimes forming alliances, have something interesting to say about disagreements involving value: contextualism, relativism and expressivism.<sup>2</sup> Additionally, there are many pragmatic, metalinguistic and hybrid accounts which offer alternative explanations of the phenomenon. In my paper, I claim that even though the existent proposals explain a lot of disagreement data by specifying various ways in which speakers may use evaluative predicates, they do not focus on providing a satisfactory explanation of the persistent auto-centric cases (Lasersohn 2004), i.e., disagreements in which each speaker utters a subjective sentence while openly and knowingly occupying his or her own perspective. To remedy that, I offer a solution which consists in supplementing the standard contextualist semantics with an explanation of this most problematic class of cases, which becomes possible due to a re-description of the phenomena using the speech act nomenclature. In section 2, I briefly discuss the problem that FD poses for contextualism. Further, I briefly talk about some accounts which aim at explaining disagreement intuitions without giving up a contextualist semantics and I claim that they do not provide a satisfactory account of persistent autocentric cases. In section 3, I argue that value terms are systematically used to perform non-assertive speech acts—praise and disapproval, which form the class of evaluations. Further, I describe a new notion of disagreement—illocutionary disagreement—and I show how it can account for FD intuitions. Additionally, I propose a characterization of evaluations as a separate kind of illocutions. In section 4, I show how my account differs from its close cousin—the conflicting attitudes view<sup>3</sup>—and argue that it provides an explanation

---

<sup>2</sup> Some absolutist views in a minimalist framework have recently been proposed too (see, e.g., Wyatt 2018).

<sup>3</sup> What I have in mind here are conflicts of non-doxastic attitudes only. Such views are typically hybrids of some account of truth-conditional content expressed in value utterances and the postulate that these utterances are connected with the existence

of disagreement intuitions, which is both more plausible and unificatory than conflict of attitudes, as it can be generalized to disagreement in general, even though it does not preclude the possibility that some affective attitudes are in fact present.

## 2. Contextualism and supplementary solutions

Few people think that value sentences are completely independent of standards or perspectives. Perhaps some speakers are inclined to believe that some deeds are absolutely and objectively morally right or wrong, but when it comes to gustatory properties such as tastiness or to aesthetic properties such as beauty, most of us seem to allow for some subjectivity or standard-dependence. When making an utterance about taste, the speaker realizes that its truth value must in some way depend on context. This context-relativity can be cashed out in different ways. One way to do it is to say that value predicates, such as *is tasty* have a hidden argument place or an unarticulated constituent which gets filled in at the context of utterance. This is what some versions of contextualism are committed to. Another way consists in keeping the content constant across contexts of utterance and assigning a truth value that is relativized to the context of evaluation, which is what relativism holds. In this paper, I am going to focus on the picture drawn by contextualism.<sup>4</sup>

According to contextualism, what is expressed in Amy's utterance of 'Brussels sprouts are tasty' in [Dialogue 1] is the proposition *that Brussels*

---

or expression of non-propositional, affective attitudes. Some of these views are called hybrid-expressivism, but since this label is not fitting for all, I will avoid using it and speak about conflicting attitudes views.

<sup>4</sup> Contextualism and relativism come in many versions, which may lead to a significant terminological confusion. Here I take only one criterion to distinguish between these two families of theories: namely, whether the value standard (for whom something is tasty or beautiful) comes into the picture as part of the proposition expressed—which I take to be a significant feature of contextualism—or whether it plays a role as one of the indices at the context of evaluation—which I take definitional for relativism. I believe that nothing in my later analysis hinges on this simplification.

*sprouts are tasty for Amy.*<sup>5, 6</sup> What is expressed by Betty, on the other hand, is the proposition *that Brussels sprouts are not tasty for Betty*. Since each speaker refers to her own standard of taste, this account of semantic content factors in the subjective character of the predicate *is tasty* and thus makes sense of the intuition of faultlessness. However, it becomes immediately clear that the proposition expressed by Betty does not contradict the one expressed by Amy and, if so, that they are not disagreeing if disagreement is construed as a pair of incoherent propositions. If we take intuition (1) for granted, however, contextualism needs to explain it anyway. It does not need to—and in most versions it does not—claim that there *is* a logical inconsistency between literally expressed propositions. Instead, it acknowledges that the appearance thereof is present and can be explained by either identifying propositional disagreement somewhere in the speaker meaning (implicature, presupposition etc.) or postulate a different construal of disagreement altogether.

Some authors (e.g., Stojanovic 2007) argue that FD is not a real problem which would require adopting a novel semantic theory (and that it does not play any significant role in adjudicating between contextualism and relativism). In each case in which FD intuitions arise, it should be possible to decide which one—intuition (1) or (2)—is symptomatic of the presence of a real phenomenon. In other words, upon inspection, it will turn out that a given case either is a real disagreement and one of the speakers is saying something false, or it's an instance of real faultlessness and the speakers are not disagreeing, but perhaps, due to a misunderstanding, they are talking past each other. The former case is not particularly contentious. It is a situation in which the speakers have the same standard in mind, but they differ about the facts of the matter. They, for example, discuss whether

---

<sup>5</sup> Here and in what follows, I ignore the time parameter, as the question of whether time should be part of the proposition is irrelevant to the aspect of the problem of disagreement I am interested in.

<sup>6</sup> Depending on the version, the standard parameter may be filled out with Amy, the group Amy belongs to, majority of people, etc. Here I focus on the 1st person perspective version of contextualism, which will take the parameter to be filled with the speaker—Amy, be it Amy simpliciter or Amy before or after brushing her teeth etc.

Pollock's *Convergence* is a beautiful painting according to most American critics, which, even though it might be a difficult question to answer in practice, *has* a definitive answer and so, one can be wrong about it. The latter case, i.e., the situation in which the speakers are truly faultless, requires a more multifaceted treatment on a case by case basis. Sometimes competent interlocutors realize they are invoking different standards, like in [Dialogue 1], and thus they are just stating something about their respective tastes, in which case the disagreement intuition should vanish (which Stojanovic refers to as the "OK/OK" situation).<sup>7</sup> In other cases, they realize that, but the intuition of disagreement does not disappear. If that happens, it makes sense to take a closer look at what else is happening in the conversation and in its context.

Some authors (Stojanovic 2007, Sundell and Plunkett 2013) notice that persistent disagreement intuitions can be symptomatic of real conflicts which have to do with coordination of practical decisions that the speakers need to make taking into account the value judgments of the other side. For instance, the crux of the argument which, on the face of it, seems like a disagreement about the gustatory value of Brussels sprouts, may actually be the question of whether they should use this vegetable in the dish they are preparing and will consume together. Another option (Barker 2013, Sundell and Plunkett 2013, Kennedy 2013) is that rather than disagreeing about whether Brussels sprouts have the property of being tasty, the speakers are negotiating a common standard of tastiness or meaning of the word 'tasty'—which makes it a metalinguistic disagreement. In other words, they are arguing about what should *count* as tasty in their common context. Yet another idea is that speakers who are disagreeing about the matters of value presuppose that they share the standard (López De Sa 2008). Once it turns out that this presupposition of commonality is false, the disagreement is revealed to be spurious. I believe that in many conversational situations these explanations of (2) are in point. However, it is conceivable that there are still going to be some cases in which the speakers would oppose all kinds of paraphrase suggested above, that is, they would not, even upon

---

<sup>7</sup> "Tarek: OK. To my taste, this ice-cream is delicious; that's all I'm saying. Inma: OK, and to my taste, it isn't delicious at all; that's all I'm saying." (Stojanovic 2007, 693)



reflection, consider their disagreement to be practical or metalinguistic. At the same time, they would believe that they occupy their own autocentric perspectives. I tend to agree both with Marques (2015) and Marques and García-Carpintero (2014) who argue that, if metalinguistic or presuppositional accounts were right, the intuition of disagreement should vanish upon enlightening the speakers. I also agree with Zeman (2016), who notices that this is the one class of cases of disputes about value that has not received a satisfactory treatment in the contextualist framework. According to him, until contextualism devises a straightforward explanation of persistent autocentric disagreement, it cannot claim it has accounted for the FD intuitions.

Finally, there are views which aim to account for disagreement intuitions by postulating conflict of non-cognitive attitudes expressed or possessed by the speakers (e.g., Buekens 2011, Clapp 2015, Huvénès 2012, Marques and García-Carpintero 2014). Such accounts often accompany contextualist accounts of the content of subjective utterances and thus can be treated as strategies supplementing contextualism with an account of FD. In section 4, I will argue that these solutions, even though plausible as accounts of value language, as accounts of FD suffer from some shortcomings which are avoided by my proposal. Before that, in the next section I present a sketch of my account, which has the potential of accounting for persistent autocentric cases of FD.

### 3. Value terms in illocutions

In order to lay the ground for my proposed explanation of what happens in autocentric cases of persistent disagreement about value, let me very briefly remind the reader of some basic notions pertaining to speech act “theory”.<sup>8</sup>

---

<sup>8</sup> It is customary to use inverted commas when talking about speech act theory, since it is not considered to be a theory with its own hypotheses, claims and so on. Rather, it is treated as a field of inquiry which, using its characteristic notions, aims at describing and investigating a large chunk of linguistic communication.

In the developed version of his framework, J. L. Austin talks about speech acts—“things done with words”—also called *illocutionary acts* or *illocutions* (1962). Austin starts with the observation that not all utterances are meant to represent how things in the world are. Plenty of what we do in conversations are other things: promising, warning, inquiring, asking, apologizing and so on. Every illocution is made up from two components: propositional content and illocutionary force. The former is—simplifying a bit—the proposition expressed, while the latter is what determines what kind of speech act is performed. Illocutions can be made with the use of performative formulae, such as ‘I promise’ or ‘I’m warning you’, but it does not need to be the case. Often the content underdetermines the force. I can warn somebody by uttering ‘There is a bull in the field’ with certain intentions and in a certain context. The proposition *that there is a bull in the field* is the content and it is uttered with the force of a warning. However, I may express the same content with the force of assertion or conjecture—thereby trying to make my interlocutor believe that it is true (or letting them know that this is what I believe). Searle and Vanderveken (1985) consider illocutions to be minimal units of human communication. They also believe that speech acts are so ubiquitous that they are the basic building blocks of the use of language: “Whenever a speaker utters a sentence in an appropriate context with certain intentions, he performs one or more illocutionary acts.” (1985, 109). It is interesting to observe what effects these acts have on our conversational and extra-linguistic reality. A successfully performed speech act puts on the speaker certain commitments concerning e.g. her or others’ future behavior as well as performance of other illocutions. If I (successfully) order somebody to  $\varphi$ , I put on them an obligation to  $\varphi$ . Moreover, if I order somebody to  $\varphi$ , I commit myself to permitting them to  $\varphi$ . This does not mean that I need to perform the act of permission, but the actions and beliefs of the persons involved are such that it is as if the permission has been granted. Speech acts, therefore, modify not only the conversational score understood as the totality of what we mutually know or believe to be true, but also our practical, normative and conversational commitments. The rules governing the connections between speech acts Searle and Vanderveken call illocutionary logic.

With this picture of communication in mind, I would like to recast some of the ideas connected with the use of evaluative language. Let us note the trivial fact that, just like many other constructions, the utterance of ‘*x* is tasty’ can be used to perform a number of speech acts. For example, it can be used to make an assertion about some fact. Doing that involves conveying the proposition that *x* is tasty to people in general, to a particular group of experiencers, to dogs, to children, or to the speaker’s 2<sup>nd</sup> floor neighbors etc. and putting it forward as true (or as something the speaker believes).<sup>9</sup> If their interlocutor knows whose standard the speaker has in mind but denies that utterance, then the disagreement is factual and propositional, not faultless. Consider, for example, the following conversation happening at the store where the speakers are discussing what kind of dog food to buy for their dog Fido:

[Dialogue 2]

Amy: Let’s get Frolic—it’s cheap and tasty.

Betty: No, it’s not tasty. Fido wouldn’t have any of that.

In [Dialogue 2] it does not seem that the speakers are expressing their personal opinions about the taste of Frolic. Rather, they are invoking the standard of taste they think their dog has. It seems, therefore, that Amy has primarily made an assertion and Betty denied it. Moreover, as we have seen, ‘*x* is tasty’ can also be used autocentrically, to convey the proposition *that x is tasty to S* (i.e. according to the speaker’s own standard). If Betty asks Amy whether she likes spinach, she might understandably assert: ‘Spinach is tasty’ just to inform Betty that spinach is tasty to her. In this context, the utterance may be meant primarily as an assertion too.

I believe, however, that in what seem like faultless disagreements, the speakers are not only making assertions. If that were the case, assuming contextualism is correct, then either denial on the part of the second speaker would be unlicensed (because they’d be denying that something is tasty for their interlocutor),<sup>10</sup> or at least one of the participants would have to be

---

<sup>9</sup> What an assertion involves is, of course, a complicated issue, which I do not need to tackle here.

<sup>10</sup> In that case it is imaginable that the denial is actually felicitous but only if the second speaker wants to argue with their predecessor about what he or she finds

mistaken about the value of the standard parameter which the other is invoking. I would like to argue that evaluative terms are systematically (although not always) used to perform a special kind of speech acts: praise and disapproval, which together can be called *evaluative speech acts* (or *evaluations*). In [Dialogue 1], under certain circumstances, Amy praises the taste of Brussels sprouts and Betty, in turn, disapproves of it. It does not just mean that Amy and Betty are saying that they praise or disapprove of the taste of Brussels sprouts. They are not reporting any mental operations. It is not *saying* something, it is *doing* something. What I am putting forward here are thus three claims:

- (1) Evaluative expressions are systematically used to perform non-assertive acts of praise and disapproval over and above expressing the proposition that something is good or bad according to one's standard.
- (2) The intuition of disagreement can be plausibly explained by invoking the conflict between illocutions (illocutionary forces)—illocutionary disagreement.
- (3) The intuition of faultlessness is accounted for thanks to the semantic content postulated by contextualism.

In a nutshell, the locution in the act of praise performed by Amy in [Dialogue 1] consists in expressing the proposition *that Brussels sprouts are tasty for Amy*. Since being tasty is a perspectival property, Amy—if sincere—has said something true and thus she has not made a mistake. The locution subsequently performed by Betty consists in expressing the proposition *that Brussels sprouts are not tasty for Betty*, which, again, is a faultless move. The conflict between the illocutions the speakers are performing—praising and disapproving—is what accounts for the disagreement intuitions. I will come back to what this conflict amounts to later in this section.

There are some straightforward reasons to consider what is happening in [Dialogue 1] as performing the special acts of praising and disapproving. One of them is that utterances like those made by Amy and Betty, as well as similar value utterances, are very naturally reported the way other

---

tasty ('No, it's not tasty to you!'). I presume that such cases constitute a minority and are not among persistent cases which I'm interested in here.

illocutions are. When someone says ‘I will do it tomorrow’, we can report what happened as ‘He said he would do it the next day’ or as ‘He promised to do it the next day’. Similarly, if somebody asks me what Amy thought about the food she had yesterday, I can either quote her words: ‘She said it was tasty’, or equally accurately report: ‘She praised it’. This suggests that describing something as tasty or beautiful can be an action over and above stating one’s own gustatory or aesthetic experience. Another quality that shows resemblance of evaluations to other kinds of speech acts is that they seem to be systematically correlated with one kind of expressions: evaluative adjectives.<sup>11</sup> At first sight there is nothing in the grammar of the sentence-type ‘x is tasty’ which would indicate that it has the illocutionary force other than assertion. However, I believe that evaluative expressions work like illocutionary force indicators<sup>12</sup>, as they are typically used, unsurprisingly, to praise or disapprove of objects and their various aspects, people, events, and actions.

Now I’m going to elucidate the notion of disagreement, which, I claim, is exemplified in such exchanges as [Dialogue 1] and propose a characterization of the types of speech acts I am postulating.

### *3.1 Illocutionary disagreement*

Many typical speech acts seem to have their opposites, i.e., acts of a similar kind but aimed at producing contrary extra-linguistic effects: assertion and denial, congratulating and condoling, forbidding and ordering (encouraging, inciting). Take for example

[Dialogue 3]

Amy (to Chris): Congratulations on your promotion! You’ll get such a nice annual bonus!

---

<sup>11</sup> Here I limit my analysis to adjectives, but possibly this account could be extended to other parts of speech (e.g., to nouns such as slurs).

<sup>12</sup> To forestall a possible misunderstanding, I do not mean to say that every use of an evaluative predicate forces us to classify a speech act as an act of praise or disapproval. Similarly, not every use of ‘I promise’ constitutes a promise (e.g., ‘If I promise to  $\phi$ , I always  $\phi$ ’). I’m grateful to an anonymous reviewer for indicating the need to clarify this point.

Betty (to Chris): No, condolences on your promotion. You'll have no more free weekends.

Here the interlocutors are performing conventional acts (*expressives*) consisting in expressing certain attitudes towards the event of Chris's promotion. These attitudes are incompatible. But the illocutionary conflict does not have to stem from attitudinal incompatibility of any kind, as [Dialogue 4] shows:

[Dialogue 4]

Amy (to Chris): I order you to do it! [alternatively: 'Do it!']

Betty (to Chris): No, I forbid you to do it! [alternatively: 'Don't do it!']

Even though what is happening between Amy and Betty does not look like a typical disagreement, there does seem to be some degree of conflict, which is indicated by the fact that both in [Dialogue 3] and [Dialogue 4] Betty's denial is felicitous.<sup>13</sup> As I have mentioned, only if we consider the acts performed in [Dialogue 1] to be something else than assertions is Betty's denial felicitous (again, assuming the contextualist picture is correct). If each speaker occupying their own perspective asserted something, denying it would have to target the assertion together with the perspective. In other words, Betty's denial would involve her asserting that Brussels sprouts are not tasty to Amy. Since this is not the case and we consider linguistic denial to be a permissible conversational move here, then perhaps it is a move in a different language game than we thought was being played.

Until now I have merely recast the notion of disagreement in terms pertaining to the speech act theoretic vocabulary. The crux of what makes praise and disapproval conflicting illocutions has to do with the way they modify the widely understood conversational score.

There is a number of accounts aimed at explaining the dynamic of a conversation. For example, Stalnaker (1978, 2002) uses the notion of the context set to model the conversational dynamics with particular focus on presupposition and assertion. The context set is defined as a set of possible

---

<sup>13</sup> Note that I am not using the term 'denial' here to refer to the act constituting opposition to assertion. By 'felicitous denial' I mean acceptable uses of expressions like 'no', 'that's not true' or even 'Nuh uh' as a reaction to somebody else's utterance.

worlds in which certain propositions are true, namely these propositions which have not yet been eliminated in the course of the conversation. Such dynamic models as Stalnaker's focus on the way assertoric speech acts restrict or otherwise modify this set. Accordingly, they concentrate on what is commonly accepted as true (or "true for the purposes of the conversation at hand"), justified or worth believing. It seems, however, that given the variety of other goals speakers want to attain in a conversation, it makes sense to postulate something like a normative dimension of the context set. Such a set would embrace values, commitments, admissible courses of action etc. and it could be used for modelling such conversations as [Dialogue 1]. Given that Amy and Betty share the context, their utterances are aimed at modifying both the "propositional" part of the context set and its "normative" part. The former is consensually updated with the propositions expressed by the speakers, while the latter is the ground on which the disagreement happens.<sup>14</sup>

Before explaining what illocutionary disagreement is in detail, I will say a few words about the way evaluations influence the common ground. In particular, I am going to sketch a picture of what commitments stemming from praise and disapproval might be and how they enter the common ground. I take the understanding of the notion of commitment from the work of Bart Geurts (2019), who proposes to analyze speech acts from a social rather than mentalist perspective.<sup>15</sup> A mentalist perspective involves understanding illocutions in terms of beliefs, intentions and other mental states that speakers want to express. That is, to make an assertion is to convey a certain belief and to make a promise is to convey a certain intention. From a social perspective, on the other hand, performing a speech act is essentially undertaking or putting on certain commitments. Geurts, along with other authors (notably Lewis 1969, Brandom 1994, Marques 2015)

---

<sup>14</sup> I choose to be non-committal with respect to the question of how the proposition enters the "normal" context set. I am, however, open to the possibility that the speakers performing evaluations simultaneously perform assertions. That would involve a version of speech-act pluralism (or rather dualism).

<sup>15</sup> Neither Geurts nor I are saying that these two perspectives are incompatible with each other. I focus on the social perspective, as it is what provides the framework which is relevant for my account of disagreement.

stresses that coordination should be the central notion guiding any viable account of communication. He says:

[T]he chief purpose of speech acts is to enable speakers to share commitments that enable them to coordinate their actions: communication is coordinated action for action coordination. (Geurts 2019, 3)

Commitments are understood in relational terms (Geurts 2019, 3):

(...) commitment is a *three-place relation between two individuals, a and b, and a propositional content, p.* (“*a is committed to b to act on p.*”)

Take, for instance, the following utterance: ‘I will clean tomorrow’. Let us say that Amy uses this sentence to make a promise to Betty and that the promise is accepted (there is appropriate uptake thereof). The commitment which thereby arises is a relation between Amy, Betty and the proposition that Amy will clean on some particular day, such that Amy will act in such a way as to make this proposition true and Betty will be entitled to expect that. There are also other related propositions for which commitments are established, e.g., that Amy will not do other things which could prevent her from cleaning etc. (Geurts 2019, 4). If the propositions turn out to be false, Amy may be held responsible for that. Commitment is thus a normative concept, in Geurts’ view. Moreover, commitments are *caused* by speech acts.

I propose to sketch the picture of what happens in the common ground when speakers perform evaluations to be somewhat along the lines of what Geurts proposes for promises. In performing the act of praise by uttering the sentence ‘Brussels sprouts are tasty’, Amy attempts to cause the relation between herself, Betty and a set of propositions, such that they both should act in accordance with their truth. These propositions are what guides coordinated action in relation to Brussels sprouts. If Amy praises Brussels sprouts and Betty does not oppose, they both will, for instance, i.a. accept the plan of having this healthy vegetable for dinner etc. *That Amy and Betty will have Brussels sprouts for dinner on some occasion* is then the proposition to which they are both committed—it enters the common ground. What enters the normative “compartment” is the relation



between Amy, Betty and this proposition (and a number of others), such that they will try to make these propositions true. On the other hand, Betty may acknowledge that her interlocutor likes the sprouts (that they are tasty to her), but she may refuse to accommodate the praise and thus to oppose taking up further commitments. In other words, she stops the relevant commitments—those tertiary relations—from being established. This is, I submit, what gives rise to disagreement intuitions in persistent autocentric cases of value disagreements and what licenses linguistic denial.

Perhaps praising the gustatory value of something does not seem to be a noteworthy update of the conversational score. Nor does it seem to deserve fighting over. The importance of common commitments becomes more striking when we think about moral evaluations. When someone uses the sentence ‘Euthanasia is evil’ to perform the act of disapproval, the context set gets significantly modified. Some acts will from then on be prohibited, some behaviors will be condemned, some plans not made and so on. Denying that euthanasia is evil—as a reaction to disapproval—constitutes a refusal to update the set of common commitments and thus gives rise to intuitions of disagreement. Disagreements about taste, even if they are less persistent or ubiquitous, can be explained by the same mechanism.

Let us now consider another example of value disagreement to see what, according to the picture I am proposing, happens when Amy performs an act of praise in a conversation with Betty:

Amy: The way Chris behaved was good.

Now, if Betty accepts the praise, we can imagine that the following can be said about their common ground from then on (the list is not exhaustive):

- what Chris did is good according to Amy’s standard;
- we will praise similar behaviors of Chris and others;
- we will reward Chris for his behavior;
- it will be understandable if Amy makes similar illocutions in the future;
- it will be admissible (*ceteris paribus*) to act like Chris did.

If Betty refuses to accept the act of praise by performing an act of disapproval, only the first of the above will be introduced to the common ground.

This brings me to a more general characterization of what illocutionary disagreement is: it is a conversational situation in which one speaker attempts to introduce the commitments (understood as relations) shared with another speaker towards certain propositions and the other blocks this attempt. The disagreement is illocutionary, since, again, both the attempt and the refusal to introduce commitments are executed by means of illocutionary acts whose constitutive role is to modify the common ground via commitments.<sup>16</sup>

### *3.2. Praise and disapproval as illocutions*

In this section I would like to discuss some reasons for distinguishing such illocutions proper as praise and disapproval, which together form the class of evaluations. Searle (1975) distinguishes the following classes: representatives (e.g., assertion), directives (e.g., ordering), commissives (e.g., promising), declarations (e.g., appointing), expressives (e.g., congratulating). Expressives seem to be the most similar to evaluations, but I believe there are reasons to keep them apart for the reasons I mention later.

According to Searle and Vanderveken (1985), the force of a type of illocution is constituted by the following seven characteristics: illocutionary point, degree of strength of the illocutionary point, mode of achievement, content conditions, preparatory conditions, sincerity conditions, degree of strength of the sincerity conditions. I will here attempt to characterize evaluations, the acts of praise and disapproval, in terms of these features.

Illocutionary point is the characteristic aim of the speech act (Green, 2020): “In general we can say that the illocutionary point of a type of illocutionary act is that purpose which is essential to its being an act of that type” (Searle and Vanderveken 1985, 112). For example, the illocutionary point of a promise is to commit oneself to doing something; that of a statement is to let others know what is the case; etc. If these main aims are not

---

<sup>16</sup> It is worth noting that illocutionary disagreement is a notion which can be applied also to assertoric speech acts. If A asserts p, which B believes to be false or unwarranted, B can block establishing the shared commitment to the truth of p. Illocutionary disagreement would then explain the “active” sense of ‘to disagree’, while propositional disagreement would explain the “state” sense (Capplén and Hawthorne, 2009).

achieved, the act cannot be considered successfully performed. The illocutionary point of evaluations is to modify the set of normative or practical commitments which are typically associated with the kind of value encoded by the evaluative expression (gustatory, moral, aesthetic etc.). This aim of evaluations is thus clearly distinct from the aim that expressives have. The illocutionary point of the latter is to simply express the psychological state of the subject with the use of some conventional devices—that is, let others know the subject has this state in performing the act.

The degree of strength of the illocutionary point allows to distinguish, for instance, conjecturing from asserting or suggesting from requesting. What they have in common is the illocutionary point, but they differ with respect to the degree of strength. Both conjecturing and asserting consist in putting something forward as true, but asserting requires a higher degree of certainty, perhaps even knowledge. The commitments they entail, even though they are of the same type, are different: one is less likely to be blamed if she conjectures something that turns out to be false than if she asserts it. Praise and disapproval also come in degrees, which are determined by the lexical meanings of the evaluative expressions used to make them. The utterances ‘This is tasty’ and ‘This is delicious’ aim at praising something, but the latter expresses the point stronger. The same goes for ‘untasty’—‘disgusting’, ‘pretty’—‘beautiful’, ‘fun’—‘amazing’ and so on.

Some illocutions need to be performed in some particular mode to be successful. For example, one can command somebody to do something only from the position of authority over that person. A promise can be made only if the promiser has some power over what is being promised (I cannot promise to you that the sun will not rise tomorrow), and the person to whom we promise something should actually want the thing promised. When it comes to praise or disapproval, it seems that no particular mode of performance can be pinned down.<sup>17</sup> There certainly are some subjects who are better-suited to making value judgments in certain situations, for example judges in competitions or experts, and presumably, these speakers

---

<sup>17</sup> It could be conjectured that to praise the taste of something successfully, one must have tasted it. However, arguably one could praise a dish by saying that it looks delicious. Such a requirement also seems to be limited to predicates of personal taste and not all evaluative expressions, so I do not want to commit myself to it.

have a higher potential of modifying the normative context set. Nevertheless, each speaker has some of that potential. Additionally, there are rules of etiquette and other social norms, which regulate the issue of whether an evaluative judgment can be made at all in a given context.

According to Searle and Vanderveken, the illocutionary force sometimes imposes conditions regulating the kind of propositional content that can be expressed in a given act. For instance, in promising or ordering, the speaker cannot express a proposition about some past event.<sup>18</sup> When it comes to evaluations as speech acts, there are some evident regularities as well as exceptions. Praising and disapproving are typically performed with the use of evaluative expressions while the standard is set to the speaker or to a group including him or her—so, in autocentric uses. For instance, praising consists in uttering, e.g., ‘x is tasty’ in which the proposition expressed contains the referent of ‘x’ and the property of being tasty to the speaker.<sup>19</sup> Nevertheless, just like other adjectives, ‘tasty’ and other evaluatives can be used referentially (‘Take the tasty one’, ‘The good guy always wins’), in which case praising may not be among the aims of the speaker. (Whether it happens anyway is a matter of further investigation.) Evaluative expressions are thus similar in function to performative expressions in that they are correlated with certain types of speech acts they are characteristically performed in. These performative expressions, according to Austin, are usually verbs in the first person present (‘I promise ...’, ‘I forgive ...’, ‘I apologize ...’, etc.); however, there are also such constructions that escape this characterization, e.g., ‘sorry about ...’, ‘welcome to ...’, ‘thank you for ...’ or ‘Hello!’. It seems, however, that it is also possible to praise someone without any reference to value. For instance, if someone asks me what I think about an online store, I may praise the owner by saying: ‘They have fresh and cheap products and they always send my order promptly’. Meaning this as praise is intelligible provided that it is common knowledge what counts as being a praiseworthy online store. The same goes for typical speech acts—

---

<sup>18</sup> Saying ‘I promise I have done what you asked me to do’ does not count as a promise in the sense relevant here. It is a non-standard use of ‘to promise’ aimed at reassuring someone that the speaker is telling the truth.

<sup>19</sup> For the sake of simplicity, I am treating propositions as structured entities, but all my claims can be recast in terms of the possible worlds theory of propositions.

even though I can apologize to someone by saying ‘I apologize’, I can attain this goal alternatively by sending them flowers.

Preparatory conditions are the conditions which have to be met for an illocution to take place—or in Austin’s terms—not to misfire. For a declarative such as ‘I hereby name you HMS Elisabeth II’ not to misfire, the speaker has to be in the position allowing him or her to christen ships. Evaluations do not seem to be constricted by many such specific conditions. Perhaps one of the few is that an evaluative expression needs to be ascribed to the right kind of things. For instance, I cannot praise the taste of someone’s behavior or the aesthetic beauty of their moral character (although I can praise the behavior and moral character themselves).

Sincerity conditions capture the requirement that in performing a given type of illocution, the speaker shall express a certain psychological state in order for it to be felicitous. In promising sincerely, the speaker expresses an intention to fulfill the promise. In apologizing, the speaker expresses remorse. This does not mean that an insincere promise is not a promise or that an insincere apology isn’t an apology. The speaker who does not have the appropriate intention has still undertaken a commitment—if other relevant conditions have been met—but the promise is not perfect. Similarly, I can perform the speech act of praising sincerely or insincerely. The former takes place when I actually express the psychological state which my words suggest I have. For example, when I sincerely praise the gustatory value of some dish, I express my non-doxastic attitude of enjoyment or liking. Note, however, that I may also just pretend that I have this attitude and insincerely praise the dish anyway to be polite, because I want to make the cook feel good or for some other reason. Nevertheless, unless I confess my insincerity, I need to face the consequences of my praise. Polite people are too often repeatedly expected to enjoy the taste of dishes they had hastily praised.

The degree of strength of the sincerity conditions is, next to the degree of strength of the illocutionary point, another parameter which allows to distinguish between similar illocutions. For instance, asking and begging are the same in all dimensions except the intensity of the psychological state expressed.

#### 4. Illocutionary account of disagreement vs. the conflicting-attitudes-views

As I mentioned above, the illocutionary account of disagreement does not fundamentally preclude some forms of the conflicting-attitudes-view (CAV).<sup>20</sup> To be precise, I do not oppose the idea that speakers typically have or communicate non-propositional attitudes in making value utterances, but I don't believe that their conflict is what is responsible for the disagreement intuitions. I will argue instead that the notion of illocutionary disagreement provides a viable alternative as a method of accounting for them.

As mentioned in section 2, one way of keeping the contextualist semantic content of value utterances while supplying means to explain the disagreement intuitions is adopting a CAV (e.g. Buekens 2011, Clapp 2015, Huvenes 2012, Marques and García-Carpintero 2014). On the face of it, this solution proposes a way of dealing with the problem of autocentric cases. A CAV, on the understanding I am invoking here, supplements the contextualist semantic content of value utterances with an account of non-propositional attitudes that are expressed, revealed or communicated alongside it. These attitudes can be desires (e.g., that the other speaker changes their taste standard or that the participants in the conversation come to share taste standards), second-order desires (e.g., that they desire to desire to change their standard), or attitudes of liking and disliking, etc. The disagreement intuition is thus cashed out in terms of conflicting attitudes.<sup>21</sup> The intuition

---

<sup>20</sup> There are a few substantially different views which have been labeled 'hybrid-expressivism', 'expressivism', or otherwise by their proponents. For the sake of simplicity and to avoid terminological inaccuracy I've chosen a different label ('conflicting-attitudes-view'). The defining feature of a CAV is the claim that value judgments are used to express, convey or otherwise communicate cognitive as well as non-cognitive attitudes (pro-attitudes). Such views are hybrid because the affective part is accompanied by some standard semantic view of truth-apt content. Since hybrid views are often proposed as a way of accounting for disagreement where it is impossible for contextualism to do so, I take their standard semantics to be contextualist.

<sup>21</sup> The FD problem is not, of course, the first time the conflict of attitudes has been proposed. For notable contributions see, e.g., Stevenson (1963).

of faultlessness is explained at the level of propositions expressed<sup>22</sup>, just as in other forms of contextualism. There are two main construals of a conflict of attitudes, described i.a. by MacFarlane (2014): nondoxastic noncotenability and preclusion of joint satisfaction. Two attitudes are noncotenable if one agent could not entertain them both at the same time. I cannot, for example, at the same time like the taste of Brussels sprouts and dislike it. Preclusion of joint satisfaction, on the other hand, envisages that two attitudes are in conflict when they cannot be jointly satisfied. This notion of disagreement can be applied to, e.g., desires. If I want to eat the last piece of cake and you want it too, both of our desires cannot be satisfied.

I do not intend to give an extensive critique of the CAV here, especially because it has already been done in other places (e.g., MacFarlane 2014). Let me just say that I agree with some authors (i.a. Marques and García-Carpintero 2014, Marques 2015, Marques 2016, Zouhar 2019, Bex-Priestley and Shemmer 2021), who argue that the notion of conflicting attitudes on its own is unable to give us a plausible explanation of disagreement intuitions in discussions about value. If we grant that the contextualist picture which envisages that in [Dialogue 1] the speakers express the enriched propositions *that Brussels sprouts are tasty to Amy* and *that Brussels sprouts are not tasty to Betty* does not explain the intuition of disagreement, it is hard to see how possessing or expressing attitudes of liking and disliking towards Brussels sprouts would do the job. Such situations would be akin to the OK/OK dialogues, as the ones mentioned by Stojanovic. It seems, therefore, that in order to do the job, the CAV needs to include an additional requirement—that the speakers strive for coordination of their attitudes, plans or norms. Such an improvement of the CAV is proposed by i.a. Marques and García-Carpintero (2014) and Marques (2015) who suggest that the attitudes in question (e.g., desires or dispositions) should be

---

<sup>22</sup> The faultlessness intuition can alternatively be construed as stemming from the fact that each speaker sincerely expresses the nondoxastic attitude they actually entertain. If I do not enjoy the taste of Brussels sprouts, then my expression of dislike can be considered faultless. Nevertheless, there are other reasons to postulate truth-conditional dimension of evaluative words next to the expressive dimension, which have to do with the problematic non-cognitive character of pure expressivism (such as the Frege-Geach problem etc.).

construed as *de nobis* higher-order attitudes. In other words, in uttering that Brussels sprouts are tasty, Amy reveals or communicates her desire that *she and her interlocutor both* desire to eat (or enjoy) Brussels sprouts. Betty's desire is for *them both* to desire not to eat (or enjoy) them.<sup>23</sup> Marques adds to that picture an explanation of why people have this kind of dispositions, according to which we generally strive for coordination for evolutionary reasons. It makes more sense in a society to share certain values:

Being disposed to eat the same sort of things enables further cooperation and altruistic behavior, and is more likely to lead to future benefits. Humans have evolved to approve of others with similar dispositions, and have evolved to disapprove of others with dissonant dispositions. Not being similarly disposed in some relevant aspects may hinder further cooperation. (Marques 2015, 9)

I sympathize with Marques' insight to a large extent. I, too, believe that often disagreements about value are driven by the need to coordinate standards and action. My two worries are, however, that (1) Marques' account does not provide a complete explanation of how the attitudes which she postulates get communicated—she is explicitly noncommittal about what kind of pragmatic mechanism is involved, which I consider to be a shortcoming of the account, and (2) that using non-propositional attitudes in explaining disagreement faces some additional difficulties (I will mention just one.) If these worries are substantiated, abandoning the concept of conflict of attitudes as crucial for explaining disagreements about value seems to be a less problematic option.

When it comes to the first point, again, I am not going to repeat arguments presented elsewhere (Hirvonen, Karczevska and Sikorski 2019), but I will just point out that the typical pragmatic mechanisms, such as presupposition or implicature do not seem to be a viable option for conveying non-cognitive attitudes or reports of speakers' having them. It should be

---

<sup>23</sup> The exact formulation of what these higher-order desires are might be different depending on the predicate, but it is irrelevant here. For details see Marques (2015) and also Marques (2016).



noted that the standard-enriched semantic content of the utterance of an evaluative sentence does not *entail* many (or even any) propositions to which the speakers become committed. It would be a long stretch to say that each of them (and perhaps more) are implicated. The fact that they enter the common ground is explainable by the constitutive role of causing commitments that speech acts have.

Regarding the second point, I believe that the illocutionary construal of disagreement fares better in accounting for disagreement *intuitions* than the one envisaged by the CAV when the attitudes in question are actually missing. After all, it may be the case that one of the speakers is insincere. For instance, Amy utters ‘Brussels sprouts are tasty’ to please a nice farmer, even though she hates the taste. Betty, on the other hand, is more straightforward. Now, the intuition of disagreement is triggered and Betty’s denial is licensed despite the fact that no conflict of attitudes arises. The illocutionary account does not require the presence of these attitudes, even though it relies on them for the evaluations to be sincere (and generally successful). The illocutionary account avoids the need to take the mentalist perspective, which, in the absence of the explanation of communicative mechanism, risks collapsing into mind-reading.

## 5. Conclusion

In this paper I have sketched a concept of disagreement consisting in performing two illocutions characterized by conflicting illocutionary forces. I have claimed that autocentric uses of evaluative predicates are often employed to perform *evaluations*—a newly-distinguished type of illocutions. I have proposed a preliminary idea of the mechanism with which illocutionary disagreement arises by employing the notion of commitment as a relation introduced to the normative part of the common ground. I suggested that when an attempt to establish a new commitment is rejected by the other party, this rejection is what accounts for the disagreement intuition. More needs to be said about this normative aspect of the common ground and the way in which commitments are related to propositions. These and related issues I leave for future research.

### Acknowledgements

I am very grateful to the Editor of this special issue Dan Zeman and to the Reviewers for their insightful comments and suggestions. I would also like to express my gratitude to the audience and organizers of the Value in Language workshop (in particular to: Graham Bex-Priestley, Bianca Cepollaro, Nils Franzén, Camil Golub, Zuzanna Jusińska, Stefano Predelli, Pekka Väyrynen, Julia Zakkou and Dan Zeman) as well as the participants of the Analytic Philosophy Department seminar at the University of Warsaw (especially to Joanna Odrowąż-Sypniewska) for their very helpful questions and remarks. Needless to say, all faults are only mine.

### References

- Austin, John Longshaw. 1962. *How to Do Things with Words*. Clarendon: Oxford, Oxford University Press.
- Barker, Chris. 2013. "Negotiating Taste." *Inquiry: An Interdisciplinary Journal of Philosophy* 56 (2-3): 240–57. <https://doi.org/10.1080/0020174X.2013.784482>
- Barker, Stephen. 2010. "Cognitive Expressivism, Faultless Disagreement, and Absolute but Non-Objective Truth." *Proceedings of the Aristotelian Society* 110 (2): 183–99. <https://doi.org/10.1111/j.1467-9264.2010.00283.x>
- Bex-Priestley, Graham and Yonatan Shemmer. 2021. "Disagreement without belief." *Metaphilosophy*. <https://doi.org/10.1111/meta.12489>
- Boghossian, Paul. 2006. "What is Relativism?" In *Truth and Realism*, edited by Patrick Greenhough, and Michael P. Lynch, 13–37. Oxford: Oxford University Press. [10.1093/acprof:oso/9780199288878.001.0001](https://doi.org/10.1093/acprof:oso/9780199288878.001.0001)
- Brandom, Robert. 1994. *Making It Explicit*. Cambridge MA: Harvard University Press.
- Buekens, Filip. 2011. "Faultless Disagreement, and the Affective-Expressive Dimension of Judgments of Taste." *Philosophia* 39 (4): 637–55. <https://doi.org/10.1007/s11406-011-9318-5>
- Cappelen, Herman, and John Hawthorne. 2009. *Relativism and Monadic Truth*. Oxford University Press. [10.1093/acprof:oso/9780199560554.001.0001](https://doi.org/10.1093/acprof:oso/9780199560554.001.0001)
- Clapp, Lenny. 2015. "A Non-Alethic Approach to Faultless Disagreement." *Dialectica* 69 (4): 517–50. <https://doi.org/10.14394/filnau.2019.0015>
- Dietz, Richard. 2008. "Epistemic Modals and Correct Disagreement." In *Relative Truth*, edited by Manuel García-Carpintero, and Max Kölbel, 239–62. Oxford: Oxford University Press. [10.1093/acprof:oso/9780199234950.001.0001](https://doi.org/10.1093/acprof:oso/9780199234950.001.0001)

- Geurts, Bart. 2019. "Communication as Commitment Sharing: Speech Acts, Communication, Common Ground." *Theoretical Linguistics* 45 (1–2): 1–30.  
<https://doi.org/10.1515/tl-2019-0001>
- Glanzberg, Michael. 2007. "Context, Content, and Relativism." *Philosophical Studies* 136 (1): 1–29. <https://doi.org/10.1007/s11098-007-9145-5>
- Gutzmann, Daniel. 2016. "If Expressivism Is Fun, Go For It!" In *Subjective Meaning: Alternatives to Relativism* edited by Cécile Meier, and Janneke van Wijbergen-Huitink, 21–46, Mouton de Gruyter.  
<https://doi.org/10.1515/9783110402001-003>
- Green, Mitchell, "Speech Acts." In *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), edited by Edward Zalta. URL = <https://plato.stanford.edu/archives/win2020/entries/speech-acts/>
- Huvenes, Torfinn Thomesen. 2012. "Varieties of Disagreement and Predicates of Taste." *Australasian Journal of Philosophy* 90 (1): 167–81.  
<https://doi.org/10.1080/00048402.2010.550305>
- Kennedy, Christopher. 2013. "Two Sources of Subjectivity: Qualitative Assessment and Dimensional Uncertainty." *Inquiry* 6 (2–3): 258–77.  
<https://doi.org/10.1080/0020174X.2013.784483>
- Kölbel, Max. 2004. "Faultless Disagreement." *Proceedings of the Aristotelian Society* (104): 55–73. <https://doi.org/10.1111/j.0066-7373.2004.00081.x>
- Lasersohn, Peter. 2005. "Context Dependence, Disagreement, and Predicates of Personal Taste." *Linguistics and Philosophy* (28): 643–86.  
<https://doi.org/10.1007/s10988-005-0596-x>
- Lewis, David. 1969. *Convention*. Cambridge MA: Harvard University Press. DOI: [10.1002/9780470693711](https://doi.org/10.1002/9780470693711)
- López De Sa, Dan. 2008. "Presuppositions of Commonality." In *Relative Truth* edited by Manuel García-Carpintero, and Max Kölbel, 297–310. Oxford: Oxford University Press. [10.1093/acprof:oso/9780199234950.001.0001](https://doi.org/10.1093/acprof:oso/9780199234950.001.0001)
- MacFarlane, John. 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press. [10.1093/acprof:oso/9780199682751.001.0001](https://doi.org/10.1093/acprof:oso/9780199682751.001.0001)
- Marques, Teresa. 2015. "Disagreeing in Context." *Frontiers in Psychology* 6 (257): 1–13. [10.3389/fpsyg.2015.00257](https://doi.org/10.3389/fpsyg.2015.00257)
- Marques, Teresa. 2016. "Aesthetic Predicates: A Hybrid Dispositional Account." *Inquiry: An Interdisciplinary Journal of Philosophy* 59 (6): 723–51.  
<https://doi.org/10.1080/0020174X.2016.1192484>
- Marques, Teresa and Manuel García-Carpintero. 2014. "Disagreement About Taste: Commonality Presuppositions and Coordination." *Australasian Journal of Philosophy* 92 (4): 701–23. <https://doi.org/10.1080/00048402.2014.922592>

- Richard, Mark. 2008. "Contextualism and Relativism." *Philosophical Studies* 119 (1/2): 215–42. <https://doi.org/10.1023/B:PHIL.0000029358.77417.df>
- Searle, John. 1962. "Meaning and Speech Acts." *Philosophical Review* (71): 423–32.
- Searle, John. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Searle, John. 1979. *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.
- Searle, John and Daniel Vanderveken. 1985. "Speech Acts and Illocutionary Logic." In *Logic, Thought and Action. Logic, Epistemology, and the Unity of Science* (volume 2), edited by Daniel Vanderveken, 109–32. Springer. [https://doi.org/10.1007/1-4020-3167-X\\_5](https://doi.org/10.1007/1-4020-3167-X_5)
- Stalnaker, Robert. 1978. "Assertion." In *Syntax and Semantics 9: Pragmatics*, edited by Peter Cole, 315–32. New York: Academic Press.
- Stalnaker, Robert. 2002. "Common Ground." *Linguistics and Philosophy* (25): 701–21. <https://doi.org/10.1023/A:1020867916902>
- Stephenson, Tamina. 2008. "Judge Dependence, Epistemic Modals, and Predicates of Personal Taste." *Linguistics and Philosophy* (30): 487–525. <https://doi.org/10.1007/s10988-008-9023-4>
- Stevenson, Charles. 1963. *Facts and Values: Studies in Ethical Analysis*. New Haven, CT: Yale University Press.
- Stojanovic, Isidora. 2007. "Talking about Taste: Disagreement, Implicit Arguments, and Relative Truth." *Linguistics and Philosophy* (30): 691–706. <https://doi.org/10.1007/s10988-008-9030-5>
- Sundell, Tim and David Plunkett. 2013. "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosophers' Imprint* 13 (23): 1–37.
- Wyatt, Jeremy. 2018. "Absolutely Tasty: An Examination of Predicates of Personal Taste and Faultless Disagreement." *Inquiry: An Interdisciplinary Journal of Philosophy* 61 (3): 252–80. <https://doi.org/10.1080/0020174X.2017.1402700>
- Zeman, Dan. 2016. "Contextualism and Disagreement about Taste." In *Subjective Meaning: Alternatives to Relativism*, edited by Cécile Meier, and Janneke van Wijnbergen-Huitink, 91–104. Mouton de Gruyter. <https://doi.org/10.1515/9783110402001-006>
- Zouhar, Marian. 2019. "On the Insufficiency of Taste Expressivism." *Filozofia Nauki* 27 (3): 5–27. <https://doi.org/10.14394/filnau.2019.0015>

# Faultless Disagreement Contextualism

Alex Davies\*

Received: 1 December 2020 / Revised March 29 2021 / Accepted: 11 May 2021

*Abstract:* It is widely assumed that the possibility of faultless disagreement is to be explained by the peculiar semantics and/or pragmatics of special kinds of linguistic construction. For instance, if A asserts “o is F” and B asserts this sentence’s denial, A and B can disagree faultlessly only if they employ the right kind of predicate as their “F”. In this paper, I present an argument against this assumption. Focusing on the special case when the expression of interest is a predicate, I present a series of examples in which the same pairs of sentences are employed, but in different contexts. In some cases, we get an impression of faultless disagreement and in some cases we don’t. I identify a pattern across these contexts and conclude that faultless disagreement is made possible, not by a special kind of predicate, but instead by a special kind of context.

*Keywords:* Disagreement; faultless disagreement; instrumental reasons; objectivity.

## 1. The subjective predicate

When is it possible for a state of disagreement to be faultless; for there to be a bona fide state of disagreement but neither party has made

---

\* University of Tartu

 <https://orcid.org/0000-0001-9978-0665>

 University of Tartu, Ülikooli 18, 50090 Tartu, Estonia.

 [alexander.stewart.davies@ut.ee](mailto:alexander.stewart.davies@ut.ee)



a mistake? Here is one widely (but not universally) assumed and tempting answer. Suppose A asserts a sentence of the form “o is F” and B asserts the (internal) negation of this sentence. It is thought that whether A and B could be in a faultless state of disagreement is decided by the kinds of predicate employed in the asserted sentences—it is the predicates employed which decide whether faultless disagreement is possible. Let’s call the predicates that many believe to be the enablers of the possibility of faultless disagreement “subjective predicates.” Different authors may believe that different sets of predicates are subjective (compare for instance Stojanovic (2019) with those she disagrees with). But typically these predicates include predicates of personal taste, aesthetic predicates and other evaluative predicates. Whichever predicates these are precisely, they are a proper subset of predicates, and the authors believe the following about them:

- (SP) It is possible for A and B to be in a state of disagreement without either being at fault only if their assertions include use of a subjective predicate.

Philosophers and linguists (e.g. Huvenes 2014, 144–45; Kennedy 2013, 259; Kölbel 2003, 21–22; Lasersohn 2005, 644; MacFarlane 2014, 2–3; Palmira 2015, 15; Solt 2018, 64; Umbach forthcoming, 2; and Wright 2001, 46–47) betray commitment to (SP) when they sort apparent disagreements into those that can be faultless, and those that cannot, more or less purely on the basis of the predicates used to express the disagreement. For instance, we might be presented with two apparent disagreements:

Office Quality

Kris:           *The Office* is funny.

Badr:           *The Office* isn't funny.

Office Author

Colleague: Ricky Gervais is the author of *The Office*.

Thomas:   Ricky Gervais is not the author of *The Office*.

Without reference being made to anything beyond the pair of assertions, it is then claimed that in Office Author either at least one of the parties is at fault or they aren’t really in a state of disagreement, yet, in contrast, this

is not so in Office Quality—as if we don't need to know anything else in order to ascertain whether either apparent disagreement can be faultless. In presenting faultless disagreements in this way, theorists of the phenomenon betray commitment to (SP). For if one cannot tell just by looking at the predicates employed whether an apparent disagreement is faultless, then one cannot distinguish Office Quality from Office Author in this way. Yet theorists think they can.

I will argue against (SP).<sup>1</sup> I don't think the possibility of faultless disagreement for exchanges of the kind between A and B has its origin in the predicates employed. I think it is a phenomenon that arises when the reasons for which A and B make their assessments of whether *o* is *F* are, in a sense to be clarified, permissive with respect to the criteria that fix what counts as *F*.

Sections 2-5 present the case against (SP). Section 6 describes some implications of the argument against (SP) for both a relativist analysis of subjective predicates and for a constructivist account of faultless disagreement that several contextualists seem to favour. Section 7 defends the argument presented in sections 2-5 against four objections.

I do not aim to convince just anybody that faultless disagreement, where it does arise, does not arise because of the peculiar properties of a special kind of predicate. There are those who deny that faultless disagreement is ever possible with any predicate (e.g. Glanzberg 2007, 16; Stojanovic 2007, 693–96). I aim to convince only those who already accept that faultless disagreement is possible at all that there is good reason to believe that it can in principle arise for any context-sensitive predicate.

I also do not mean to deny, in what follows, that there are differences between predicates which can be described with the words “subjective” and “objective.” I particularly have in mind the capacity of some adjectives (possibly different ones in each case) but not others: to appear in the complement position of “find”; to felicitously combine with an explicit experimenter argument; and to require that the speaker have the right kind of first-

---

<sup>1</sup> Zakkou (2019, 17) explicitly acknowledges that judgements about whether a given pair of assertions constitutes a case of faultless disagreement are context-sensitive. But this observation is made in defence of the possibility of faultless disagreement altogether rather than as an observation which might threaten (SP).

hand experience. I am not going to try and convince anyone that if you put a predicate in the right context, all this can be had. I am concerned only to deny that a special kind of predicate is a necessary enabling condition of, specifically, faultless disagreement.

## **2. Reasons for making an assessment: the same, different and the same again**

I want to draw your attention to an interplay between the reasons for which A and B engage in an assessment of whether something is F and our intuitive judgements about whether A and B disagree with each other, and if they do, whether their disagreement is one in which somebody has to be mistaken. We'll begin with, and will subsequently riff upon, an example from Richard (2004):

Didi (in her context C1): Mary's rich.

Naomi (in her context C2): Mary is not rich at all.

Richard defines C1 and C2 as follows:

### *Example 1*

Suppose, to take an example, that Mary wins a million dollar lottery. Didi is impressed, and remarks to a friend 'Mary's rich.' Naomi, for whom a million dollars is not really all that much, remarks in a conversation disjoint from Didi's, 'Mary is not rich at all'... Suppose that there is no difference between the two conversations in the point of assessing people as rich or otherwise. (Each conversation began with the observation that some wealthy person doesn't deserve to be rich, and each of the women is now idly assessing people as rich or otherwise, and then assessing whether the rich ones deserve their wealth.) (Richard 2004, 218)

Looking at this, it seems that Didi and Naomi disagree: Didi thinks that Mary is rich whereas Naomi thinks that Mary is not rich. Moreover, it doesn't seem that one of them must be mistaken in making the assertion that she does. In other words, there's not much to distinguish this pair of assertions from Office Quality. It seems to be a case of faultless disagreement.



Let's now consider two variations on example 1.

*Example 2*

Suppose that Mary wins a million-dollar lottery. Didi is talking with a good friend, of modest income, who's just suffered a burglary: all the children's Christmas presents were stolen, and they don't have the money to buy any replacements. Didi, hearing this sob story, remembers Mary. "Hey, Mary's got a good heart. And she's just won the lottery—literally. Mary's rich. I'm sure she'd help you guys out." Naomi, herself someone for whom a million dollars is not a lot of money, has a friend who is looking for someone who might buy his apartment: worth 98 million dollars. Her friend remembers that Mary was looking to buy a new apartment and asks Naomi to inquire into whether Mary would be interested. But Naomi replies, "Mary? Mary's not rich. You're barking up the wrong tree with her."

Here, we once again get faultlessness: it doesn't seem that either of Didi or Naomi is making a mistake in her assertion—they're both right. However, in marked contrast to Richard's original example 1, it's pretty clear that here Didi and Naomi are not in a state of disagreement with each other.

Finally, consider this:

*Example 3*

Suppose that Mary wins a million-dollar lottery. Didi is talking with a good friend, of modest income, who's just suffered a burglary: all the children's Christmas presents were stolen, and they don't have the money to buy any replacements. Didi, hearing this sob story, remembers Mary. "Hey, Mary's got a good heart. And she's just won the lottery—literally. Mary's rich. I'm sure she'd help you guys out." In a different conversation in another part of the city, Naomi hears the same sob story and is asked whether she knows anyone who might be kind-hearted and wealthy enough to help out the family. She's asked specifically about Mary in this regard. Naomi replies, "Mary? Mary's not rich. You're barking up the wrong tree with her."

Here, we get the reverse of what we had in example 2. We get disagreement—just as we had in example 1—but we've lost faultlessness: Naomi is mistaken—she's saying something false.

In all three examples we have two assertions, each assertion taking place in a different conversation from the other in the pair. But in example 1, we seem to get both a state of disagreement and the sense that neither party to the disagreement is mistaken. In example 2, we retain faultlessness but we lose disagreement. And in example 3, we retain disagreement but we lose faultlessness. So what's making the difference?

### 3. Permissive reasons for making an assessment

Instrumental reasons are the kind of reason you give for doing something when you say it is a means to some further end (Kolodny 2018). For instance, my reason for walking into town is to buy milk. My reason for driving right now is that I'm going to Watford. Notice that if you're doing one thing for the reason that you're doing something else, then that something else often introduces constraints on how you should do that thing. If you're driving to Watford, then when you approach a particular T-junction, it may well be the case that you ought to turn a certain direction, because that's the way to Watford and the other direction is not. Thus, the reason for which you're driving introduces normative constraints upon *how* you drive. The same is true of an assessment of whether *o* is *F* and the reasons for which you are making the assessment. Sometimes the reason for which you're making an assessment of whether *o* is *F* introduces normative constraints on the criteria you ought to employ for what counts as *F*, just as the reason for which you're driving introduces normative constraints on how you drive.

With this in mind, think back to the examples just provided. For what reasons are Didi and Naomi making assessments of whether Mary is rich? Answer: for different reasons in the different examples. In example 1, they are both making their assessments "idly" as the basis for subsequent discussion about whether rich people deserve their money. If the assessments are idle, this presumably means there's nothing about the reasons for which they make their assessments which could be appealed to in defending either the inclusion of Mary in, or the exclusion of Mary from, the set of rich people. In example 2, things are different. Here neither Didi nor Naomi is idly making her assessment: the reason why Didi is making the assessment

is in search of someone who could help out her friend and her friend's family for Christmas. The reason why Naomi is making the assessment is in search of someone who has wealth enough to purchase a 98 million-dollar apartment. Intriguingly, in this example, their reasons *do* provide something which could be appealed to in defending either the inclusion of Mary in, or the exclusion of Mary from, the set of rich people. With respect to Didi's reason for making her assessment: all that matters for achieving Didi's overarching goal is that someone be identified who has ample disposable savings to spare some money to buy some Christmas presents for a family of modest income. Given that in making her assessment, this was all Didi was doing, and given that Mary has suddenly got an excess of one million dollars, Mary fits that bill. For this purpose, Mary should be classified as rich. With respect to Naomi's reason for making her assessment: all that matters for Naomi's overarching goal is that someone have enough disposable income to afford a 98 million-dollar apartment. Mary does not fit this bill. For this purpose, Mary should not be classified as rich. So in example 2, we have *divergent* purposes and each purpose places normative constraints on whether Mary should be considered rich (for the respective purpose), which were absent in example 1. In example 3, with respect to the reasons why Didi and Naomi are making their respective assessments, things are different once again. This time, just like in example 1, each of Didi and Naomi is making her assessment for the same reason but the reason isn't the same as in example 1. This time it's the reason Didi was making her assessment in example 2: i.e. in search of someone who could help out the unfortunate friend and her family for Christmas. Just as in example 2, if the assessment is being made for this reason, then Mary ought to be counted as rich. This has the result that when Naomi denies that Mary is rich, she seems to be incorrect, given the reason for which she's making the assessment of whether Mary is rich.

There are two variables I think it's worth keeping track of across these examples: whether there is a divergence of reasons for assessment between Didi and Naomi; and whether the reason for the assessment is the kind of thing that can be appealed to in defence of a particular criterion for being rich. Let's give a name to this kind of reason. A "permissive reason" is a reason for which you perform an assessment of whether *o* is *F* that satisfies the following condition:

*Permissive reason*

A reason *r* for making an assessment of whether *o* is *F* is permissive with respect to two criteria for deciding whether *o* is *F* if *r* permits those two criteria and according to one, *o* won't count as *F*, but according to the other, *o* will count as *F*.

Reason *r* permits a criterion if it doesn't require (in the way that instrumental reasons can require things of that for which they are a reason) that one use another criterion.

In the examples provided, the interaction between these two variables and the felt presence of faultless disagreement can be described with the following table:

	Both permissive	Both not-permissive	One permissive, one not-permissive
Same reason	Faultlessness Disagreement (Example 1)	No faultlessness Disagreement (Example 3)	NA
Different reason	?	Faultlessness No disagreement (Example 2)	?

Table 1

Let's now try to fill in the two question-marks. We can do that by considering two more examples, once again, riffing on Richard's original.

*Example 4*

Didi's in a conversation that began with the observation that some wealthy person doesn't deserve to be rich, and Didi and her friend are now idly assessing people as rich or otherwise, and then assessing whether the rich ones deserve their wealth. Suppose that Mary wins a million-dollar lottery. Didi is impressed, and, in the course of this conversation, she remarks to her friend 'Mary's rich.' In another part of town, a friend of Naomi's has expressed an interest in meeting someone rich, just to know what it feels like. The friend suggests Mary as an

option—with her newfound wealth. Naomi, for whom a million dollars isn't a lot of money, rejects the idea, "Oh no. Mary's not rich."

Do Didi and Naomi disagree? I think they do: were they to meet, you couldn't appeal to their differing reasons for making their assessments to demonstrate that they don't really disagree. And I also think there's nothing wrong with either assertion. In this example, we have two *different* reasons why assessments are being made of whether Mary is rich. Yet in contrast with example 2, both of these reasons could reasonably be described as "permissive": neither provides us with a basis upon which we could defend a particular stance on how to classify Mary *vis a vis* being rich: the reasons don't favour one criterion over another.

Look now at a final variation on Richard's example 1:

*Example 5*

Didi's in a conversation that began with the observation that some wealthy person doesn't deserve to be rich, and Didi and her friend are now idly assessing people as rich or otherwise, and then assessing whether the rich ones deserve their wealth. Suppose that Mary wins a million-dollar lottery. Didi is impressed, and, in the course of this conversation, remarks to her friend 'Mary's rich.' In another part of town, Naomi, herself someone for whom a million dollars is not a lot of money, has a friend who is looking for someone who might buy his apartment: worth 98 million dollars. Her friend remembers that Mary was looking to buy a new apartment and asks Naomi to inquire into whether Mary would be interested. But Naomi replies, "Mary? Mary's not rich. You're barking up the wrong tree with her."

My impression here is that although each of Didi and Naomi is making no mistake in making the assertion that she does, they are not in a state of disagreement with each other. In this example, we have a divergence of reasons for which assessments are being made, and one of these reasons is permissive (Didi's) whereas the other reason (Naomi's) is not.

These two examples allow us to complete the table:

	Both permissive	Both not-permissive	One permissive, one not-permissive
Same reason	Faultlessness	No faultlessness	
	Disagreement (Example 1)	Disagreement (Example 3)	NA
Different reason	Faultlessness	Faultlessness	Faultlessness
	Disagreement (Example 4)	No disagreement (Example 2)	No disagreement. (Example 5)

Table 2

At this point, it becomes tempting to formulate a hypothesis on the basis of the pattern witnessed in the table. At the start of this paper, we considered the generic situation in which A asserts a sentence of the form “o is F” and B its negation. A and B are each making an assessment of whether o is F but coming down on different sides of the issue. The pattern seen in the table suggests that A and B can be in a state of faultless disagreement only if the reasons for which they make their respective assessments are both permissive with respect to how to assess whether o is F—where this does not mean they have to be making their assessments for the same reason (see example 4).

#### 4. Predicates of personal taste and non-permissive reasons for making an assessment

If this hypothesis is correct, then we should expect it to apply to assertions made using paradigmatically subjective predicates as well. In particular, we should expect to be able to block an impression of faultless disagreement by appropriate modification of the reasons for which asserters are making their respective assessments of whether some o is F. Let’s see if we can do that:

##### *Example 6*

Maksim’s job is to visit stand-up comedians doing a gig and find out whether their kind of humour would suit the audience of the Comedy

Cellar, where his boss works. To do this, Maksim broadcasts the gig to a sample of the Comedy Cellar audience. Members of the sample can then give an indication of whether they think the comedian is funny. Maksim watches a comedian who calls herself Sergeant Knock Knock. The sample listens. The sample is unimpressed. Maksim's boss calls Maksim to ask for the verdict. "So, is she funny?" "No, she's not funny. We better go with that other one from last week." Nikolai is at the same gig doing the same thing—but Nikolai works for the Comedy Penthouse, a different club. It's well known that their clientele have a different sense of humour to those folks down at the Comedy Cellar. Nikolai's boss calls him up, and asks for the verdict, "So, is she funny?" "Yes, she is funny. We should get her up to the Comedy Penthouse next week if we can."

It seems to me that neither Maksim nor Nikolai are at fault in making their respective assertions. However, they aren't in a state of disagreement with each other. This is as we would expect because example 6 is modelled on example 2. But things change if we approximate the kind of situation we see in example 1:

*Example 7*

Nikolai heads on over to his local, and Maksim over to his, where each meets with his respective gang of friends. Nikolai's friends ask (just to break the ice as Nikolai arrives) how the act was. Thinking how he personally felt about Sergeant Knock Knock, Nikolai replies, "She wasn't funny. But the Comedy Penthouse audience liked her. So we'll be seeing more of her." Maksim's friends ask him the same question, and for the same idle, ice-breaking reason. Maksim replies, "She was funny. It's a shame the Comedy Cellar's audience didn't agree."

Here it seems that Nikolai and Maksim do disagree and that neither of them need be mistaken in his assertion: as we would expect, given the parallel with example 1.

We could construct further parallels with examples 3-5. For instance, we could adapt example 6 so that it parallels example 3 by stipulating that Maksim and Nikolai work for the same comedy club. But for reasons of space, I won't do this. It should nonetheless be clear that an impression of faultless disagreement comes and goes with changes of context just as

much for assertions that deploy a paradigmatically subjective predicate (“is funny”) as for assertions that deploy another kind of predicate (“is rich”).

## 5. Rejecting (SP)

I conclude on the basis of these examples that (SP) is false. Whilst holding fixed the expressions employed in a pair of assertions, and whilst changing the contexts in which the assertions are made, we witness shifts in impressions of faultless disagreement. What seems to be important to the presence of an impression of faultless disagreement is that the assessments of whether *o* is *F* which are being made with each assertion are being made for reasons that are permissive with respect to two criteria for whether *o* is *F*, such that on one, *o* counts as *F* but on the other it does not. If we have that, then we get faultless disagreement. If not, then we don't. It seems then that faultless disagreement arises when the broader non-communicative goals of the relevant pair of asserters fail to place sufficient instrumental constraints on the criteria employed for deciding whether some item falls into a given category: the rich, the funny. This absence of constraint is not something that can happen only for a special class of “subjective” predicates.

## 6. Some consequences of rejecting (SP)

The position I'm putting forward here about when faultless disagreement is possible has implications for both contextualist and relativist analyses of the meanings of subjective predicates. Assume that the content of a sentence is a function from indexes to truth-values. The relativist thinks that whereas predicates of personal taste have a built-in, context-invariant metasemantics which lets a context of assessment (rather than of assertion) set a parameter in the index, other predicates have a built-in, context-invariant metasemantics which lets only the context of assertion (rather than of assessment) set the parameters in the index (e.g. MacFarlane 2014). This predicate-based analysis is adopted because the relativist thinks that the



possibility of faultless disagreement is to be explained by the kind of predicate employed. But the examples presented above suggest that this is a mistake. If a relativist analysis is the best way to account for the possibility of faultless disagreement, then the metasemantics itself should be context-sensitive in the following respect: there is variation across contexts of assertion in whether a sentence containing a predicate has its judge index set by the context of assertion (and so doesn't permit faultless disagreement) or instead by a context of assessment (and so does permit faultless disagreement). It shouldn't be a context-invariant feature of the predicate that it is one or the other.

Contextualists understand the faultlessness of faultless disagreement as arising from a divergence of contents for the predicate employed in the two relevant assertions (a divergence that makes possible the consistency of the sentences containing these predicates). Different contextualists adopt different accounts of the impression of disagreement. For instance, Zakkou (2019) proposes that the impression of disagreement arises because asserters are pragmatically conveying propositions to the effect that their own criterion for determining what is F is the best: propositions which are inconsistent. Others understand the impression of disagreement as arising out of a practical conflict of some sort. For instance, Barker (2013) and Sundell (2011) think that the impression of disagreement reflects a disagreement about how an expression should be employed, whereas Marques' (2014, 2016) and Marques and García-Carpintero's (2014) think the impression of disagreement reflects asserters' possession of desires about what asserters desire, where these higher order (*de nobis*) desires cannot be jointly satisfied. The position defended in this paper has implications for at least some of these views (just as it does for the relativists). I have space to discuss just one of these views.

Although I want to be a little cautious in ascribing this view, it seems at least in places that Barker (2013, 247–49) and Sundell (2016, 808–17) (as well as Kennedy 2013, 274 and Khoo and Knobe 2018, 21–27) are constructivists about the content of context-sensitive expressions: they think that such expressions have the contents they do because there is an agreement between their users that these expressions have these contents. So when this agreement falls away (as in a metalinguistic dispute about

how the expressions ought to be used), there is no fact of the matter about the content of the relevant expression. Like the position defended in this paper, constructivism implies that faultless disagreement is in principle possible almost anywhere (or at least, wherever context-sensitive language is employed). But in contrast with the position here defended, constructivism implies that a metalinguistic disagreement creates faultlessness. This is because, given constructivism, such disagreement just is the absence of what makes fault possible (i.e., metalinguistic agreement and the resultant linguistic fact). The observations made in this paper suggest that this isn't quite right. When the reasons for which the assessments are being made are not permissive, a metalinguistic disagreement can only be an exploration of linguistic facts that are already there (rather than yet to be constructed through the achievement of agreement). If, for instance, the reason for which assessments are being made require that persons with (only) one million dollars of spare cash be classified as below the threshold for being rich (as in example 3), then (if we understand faultlessness along contextualist lines) that's part of the content of "is rich" in the context. If a disagreement breaks out about this, the disputants will be exploring a pre-existing normative landscape—one put in place by the reasons for which they are making their assessments of whether someone is rich (metalinguistic disagreement is an exploration of the normative implications of pre-existing practical commitments). Consequently, the disagreement will be faultless only if the reasons for which assessments are being made are permissive (if not, we have a situation like example 2 or example 3, wherein either there's no disagreement or someone is making a mistake). Sundell seems to suggest something incompatible with this when he explains why persons would ever be motivated to engage in a metalinguistic disagreement: to convince his readers of this, he points to just those circumstances in which the reasons why assessments are made are *not* permissive. If the position defended in this paper is correct, such contexts won't make disagreements faultless (because such contexts resemble either example 2 or example 3). The contexts capable of giving rise to faultless (on the contextualist analysis, metalinguistic) disagreements are those in which disputants are not pursuing the kind of demanding non-communicative tasks Sundell envisages. They're instead

engaged in tasks that don't undermine the impression of disagreement but which also fail to give language the rigour needed to make anyone in the disagreement mistaken.<sup>2</sup>

Of course, this raises a question. Sundell explains why two people would be motivated to express and pursue a metalinguistic disagreement because of the consequences that depend upon the content with which the relevant predicate is employed. But if faultless disagreement arises in a context where nothing much hangs on with which content a predicate is used, then what motivates this behaviour? Well, there are motivations that don't depend upon there being any weighty contextual consequences to the choice of content to assign a context-sensitive word. Argument can be used simply as a form of sociability: done for itself, for kicks (Schiffrin, 1984). It can be used to do identity display, and show off who one is (Davies, forthcoming). It can be used to make fun of someone ("That's not a knife..."). It might even be that the disputants do not know that they're in such a context, and so use argument to find that out for themselves whether one content is more suited to their purposes than the other, and for it to be difficult to discern that one is no worse than the other.

## 7. Objections and replies

I turn now to four objections that might be raised against the hypothesis that faultless disagreement is made possible when assessments are being made for permissive reasons (rather than because a special kind of predicate is being employed).

---

<sup>2</sup> The difference between constructivism and the position defended in this paper parallels a broader debate between those who favour a metasemantics of context-sensitive expressions that places greatest emphasis upon the agreement of the users of those expressions (e.g. King 2013 and Michaelson 2014), and those who favour a metasemantics that places greatest emphasis upon the non-linguistic actions in which the use of language is embedded (e.g. Davis 2018 and Dobler 2019), as the source of what makes a given expression have the content it does in a given context.

### 7.1 *But that's just "rich"*

One concern is that examples 1-5 all involve the same predicate “is rich” and one might worry that what goes for this predicate doesn’t go for others. I don’t have the space to reproduce the same kind of examples that I have given above for other predicates. Doing this would also take space away from other things I should be doing here. But other examples are easy enough to construct on the template provided by examples 1-5. We find a predicate “is F/are F / Fs / F-ed” and we find some object we might apply this predicate to (where the object could be a particular or a kind), where the predicate and the object are such that, the predicate denotes a property such that we can think of two criteria which establish whether an object has that property such that on one criterion, the object doesn’t have the property but on the other criterion it does. We then build the different sorts of context found in table 2. For example, take Office Author from the opening of this paper. What is required to be the author of something? Here are two ways to think of that. On the first, Gervais is the author of *The Office* only if he has authority over its interpretation: Gervais’ intentions behind the scripts determine the content of the script. On the second, Barthesian way to think of authorship, Gervais is just a “scriptor”, a person who creates the scripts, but who’s intentions do not have any authoritative role in deciding what the content of the script is. We could imagine that there is a pair of people, one asserts “Gervais is the author *The Office*” and one asserts that sentence’s negation, and in asserting what she does, each has a different view about what is required to be the author of something. And we can imagine these assessments of whether Gervais is the author being made by each asserter through her respective assertion being done for various different reasons. Sure enough, if, for instance, they are each speaking for idle reasons, then we’ll have a situation much like we see in example 1; one in which there’s an impression of faultless disagreement. If on the other hand, each makes her assessment for a non-permissive reason, then we’ll either lose the impression of disagreement (as in example 2) or we’ll lose the impression of faultlessness (as in example 3), depending upon further details about their respective non-permissive reasons for making their respective assessments.

### *7.2 Elaboration objections*

Several have presented the following objection to the possibility of what I'm describing (Hawthorne 2004, 104; Marques & García Carpintero 2014, 712–14; Schaffer 2011, 213; Stanley 2005, 55–56). The objection is that when we allow further elaboration on what each asserter meant by her assertion, then the sense that they disagree vanishes. For instance, consider the following two assertions:

*Example 8*

A (assuming Yasser is 1.96 metres tall, discussing the height of basketball players): Yasser is short.

C (responding to A, and aware that the height of basketball players is discussed, but who has a different very precise perspective on how to draw the line of height for players, which has led him to draw the line for shortness at 1.956 metres assuming the same about Yasser as A): Yasser is not short.

At first glance, this looks like a case of disagreement. A thinks that Yasser is short because A thinks that to be not short you need to be taller than 1.96 metres. C thinks that Yasser is not short because C thinks that if you are taller than 1.956 then you are not short. But, what if the following elaboration is offered by A?:

A: That does not contradict what I said; I was just saying that Yasser is short for a basketball player on rough estimates for the purposes of coffee talk. I was not contemplating your own estimate; thus I was not wrong. (Marques & García Carpintero 2014, 712–14)

Sure enough, if this elaboration is offered (and true), then, I concur, the impression of disagreement between the two assertors vanishes. But notice that in this elaboration a distinction of the reasons why the two assertions are being made is drawn: A was making an assessment for 'purposes of coffee talk' whereas C was making an assessment presumably for other purposes—or else there would be little point in A drawing attention to her own. In other words, the elaboration offered as evidence against Richard makes the whole scenario better resemble example 2 than Richard's original example 1 (For further discussion see Davies 2017).

In response to this, one might object to the purported importance of the *reasons for which* each asserter is making an assessment of Yasser's shortness. One might think that for an elaboration to dislodge the impression of disagreement, it suffices to point out that each asserter draws the line for being short differently. That in and of itself suffices to show that they don't disagree about whether Yasser is short: appeals to relevant differences in the reasons for which each asserter is making her assessment are unnecessary.

But surely this is an exaggeration. Two people can perfectly well be in a state of disagreement about whether *o* is *F* because (not despite the fact that) they disagree about what is required or sufficient for something to be *F*. The Russian government thinks that Estonia freely voted itself into the Soviet Union. The Estonian government disagrees: Estonia was occupied by the Soviet Union, it did not freely vote itself into the Soviet Union. The Russian and Estonian governments very likely have different understandings of what is required for Estonia to have freely voted itself into the Soviet Union. That doesn't mean the governments don't disagree about whether Estonia freely voted itself into the Soviet Union. Differences of this sort just don't have the capacity, in and of themselves, to undermine the presence of a state of disagreement. Likewise, the mere fact, in and of itself, that *A* and *C* think that shortness is marked off at different points on a scale of height doesn't show that they don't disagree about whether Yasser is short. If the impression of disagreement is to vanish, it is important to distinguish different reasons why the assessments of Yasser's height were being, with it being clear that the reasons for making each assessment favour the respective assertions. If that isn't clear, then it might become unclear whether we have in hand an example which most closely resembles any one of our earlier examples 1-5. But that unclarity wouldn't show that there are no examples like 1.

Anyone who accepts that faultless disagreements are possible and that the best account of them makes the faultlessness *veritic* must accept that a mere difference in the way persons draw the extensional boundary for what is *F* doesn't suffice to disqualify them from being in a state of disagreement with one another. It would be incoherent to then insist that such differences *do* suffice to thwart disagreement.

### 7.3 Comparative predicates

It has been claimed that faultless disagreement may be possible with an objective predicate like “tall” but not with the same adjective’s comparative form and that, in this respect, the adjective itself “tall” seems to differ from, for example, the adjective “fun” (Kennedy 2013, 269; Solt 2018, 60; and Umbach forthcoming, 5). For example, consider the following pair of pairs of assertions provided by Solt (2018, 60):

*Example 9*

Speaker A: The chili is tastier than the soup!

Speaker B: No, the soup is tastier!

*Example 10*

Speaker A: Anna is taller than Zoe.

Speaker B: No, Zoe is the taller of the two!

Solt classifies example 9 as a faultless disagreement. She classifies example 11 as factual only. And just looking at these as they stand, ignoring what more may be going on in contexts of each pair, I agree that the pair of pairs of assertions give rise to the impression Solt describes. However, I nonetheless think it possible for an impression of faultless disagreement to arise when the relevant pair of assertions includes use of the comparative form of “tall”. What’s difficult about thinking of ways in which this could happen is that whether one item is taller than another is not something that could be open to dispute by two reasonable people who each have made no mistake about how tall the two items are.<sup>3</sup> But if we can find ways in which different scales of height would result in different orderings of the two items then, it seems, there’ll be room for two people to disagree about which is taller than the other without either having made a mistake (provided that the reasons for which each makes her assertion are permissive with respect to the categorization of these two items in terms of their relative heights). Consider the following example:

---

<sup>3</sup> “Taller than” seems not to be multidimensional (Kennedy, 2013). My point here is that it’s not impossible for “taller than” to be multidimensional.

*Example 11*

During the course of a day, in some sense, your height changes. This happens because the spinal discs between your vertebrae are largely made of water. When you place weight on the disks in your back (as when you are standing or sitting upright), they compress (over the course of several hours). But when you're lying down, they expand. Ana, Bea, Cat and Dee are four competitive girls. They compete over everything: who can fit the most gob-stoppers in her mouth, who can jump the furthest etc. Suppose that Ana recently broke a leg and for this reason spends most of each day lying down. When they stand against a wall, Ana's height is a centimetre or so higher than Cat's. Suppose that if Ana and Cat had been lying down all day, then Cat would be taller than Ana. Being competitive girls, each girl wants to be taller than the other. For that reason, Cat adopts the view that the proper way to compare two persons' heights is to ensure that they have both been lying down for the same amount of time prior to measurement. For the same reason, Ana adopts the view that the proper way to compare two persons' heights is to stand them against a wall and measure their heights—regardless of whether one of them has been lying down recently. Finally, suppose that Bea is a friend of Ana's who agrees with Ana about the proper way to measure relative height and Dee is a friend of Cat's who agrees with Cat about the proper way to measure relative heights. Now, suppose that in one context, Ana and Cat are talking about whether Ana or Cat is taller, and in another context, Bea and Dee are talking about whether Ana or Cat is taller. In each context, there's nothing but banter going on: these are permissive contexts. Nothing in either context settles how relative height is to be understood in that context. In the first context, Ana says, "I am taller than you," and in the second context, Dee says, "Ana is not taller than Cat." (Davies 2017, 871)

Ana and Dee are in a state of disagreement. But neither need be mistaken in her assessment. The difference in impression we get between examples 9 and 10 is not a context-invariant effect.



#### 7.4 Experimentally demonstrated differences

Experimental findings indicate that if a person is presented with a pair of assertions, whether she will think that faultless disagreement is possible for that pair of assertions will be affected by the predicate employed in the pair; i.e. by whether it is paradigmatically subjective or not (see Cova & Pain 2012 and Solt 2018, 65). If the reasons for which assessments are being made by the asserters are not provided, then, it might seem, we should predict that there would be no difference in reaction to sentences containing different kinds of predicate. Since there is, this speaks against the view that the reasons for which assessments are made play a crucial role in determining where impressions of faultless disagreement arise.

But I don't think we're forced to accept this conclusion. When we hear certain combinations of words, we are more likely to associate the combination with one kind of context than another. Hear the words "would you like fries with that?" and we think of a cashier at a fast-food restaurant inviting us to expand our initial order. Hear the words "tuck your shirt in!" and we imagine a teacher ordering around a schoolchild. But these sentences don't need to be used in only those contexts that come first and most strongly to mind when we hear them out of the blue. Words make certain contexts salient. But this doesn't mean that the features the words have when thought of as used in a most salient context are context-insensitive properties of the words themselves.

Given this, it is even to be expected that when presented out of the blue with a pair of assertions of sentences that include, for instance, predicates of personal taste, we will be inclined to imagine their use in a salient kind of context, whereas when presented with a pair of assertions made using a kind of predicate typically classified as "objective", we are inclined to imagine their use in a (different) salient kind of context. If the most salient contexts of use differ in whether the reasons for the assessments made with the respective assertions are permissive in the right way, then we'd expect what's witnessed by Cova and Pain, and by Solt. But this wouldn't count against the position being defended in this paper.

On the contrary, the position defended in this paper implies that there is a potential confound in these studies' designs, insofar as these studies are used to derive conclusions about predicates *per se*. If one is interested in

where and why faultless disagreement seems possible, and if context makes a difference to this in the way seen in our examples 1-7, then that factor should be controlled for. To date, they haven't been. It's just been supposed that if no context is provided, then context plays no role in patterns of impressions of faultless disagreement.

### Acknowledgements

I'd like to thank the participants of the University of Tartu Theoretical Philosophy WiP seminar, the referees for *Organon F* and the editor of this issue for pushing me to improve this paper far beyond the state of the original submission. The research that led to this paper was supported by the European Union's Regional Development Fund through the Centre of Excellence in Estonian Studies.

### References

- Barker, Chris. 2013. "Negotiating Taste." *Inquiry* 56 (2-3): 240–57. [doi.org/10.1080/0020174X.2013.784482](https://doi.org/10.1080/0020174X.2013.784482)
- Cova, Florian, and Nicolas Pain. 2012. "Can Folk Aesthetics Ground Aesthetic Realism?" *The Monist* 95 (2): 241–63. Retrieved from <https://www.jstor.org/stable/41419025>
- Davies, Alex. 2017. "Elaboration and Intuitions of Disagreement." *Philosophical Studies* 174 (4): 861–75. DOI: [10.1007/s11098-016-0710-7](https://doi.org/10.1007/s11098-016-0710-7)
- Davies, Alex. 2018. "Communicating by Doing Something Else." In *The Philosophy of Charles Travis*, edited by Tamara Dobler, and John Collins, 135–54. Oxford: Oxford University Press.
- Davies, Alex. forthcoming. "Identity Display: Another Motivation for Metalinguistic Disagreement." *Inquiry*. <https://doi.org/10.1080/0020174X.2020.1712229>
- Dobler, Tamara. 2019. "Occasion-Sensitive Semantics for Objective Predicates." *Linguistics and Philosophy* 42 (5): 451–74. <https://doi.org/10.1007/s10988-018-9255-x>
- Glanzberg, Michael. 2007. "Context, Content and Relativism." *Philosophical Studies* 136 (1): 1–29. [doi.org/10.1007/s11098-007-9145-5](https://doi.org/10.1007/s11098-007-9145-5).
- Hawthorne, John. 2004. *Knowledge and Lotteries*. Oxford: Clarendon Press.
- Huvenes, Torfinn Thomesen. 2014. "Disagreement without Error." *Erkenntnis* 79 (1): 143–54. [doi.org/10.1007/s10670-013-9449-0](https://doi.org/10.1007/s10670-013-9449-0)

- Kennedy, Christopher. 2013. "Two Sources of Subjectivity: Qualitative Assessment and Dimensional Uncertainty." *Inquiry* 56 (2-3): 258–77.  
<https://doi.org/10.1080/0020174X.2013.784483>
- Khoo, Justin, and Joshua Knobe. 2018. "Moral Disagreement and Moral Semantics." *Nous* 52 (1): 109–43. <https://doi.org/10.1111/nous.12151>
- King, Jeffrey C. 2013. "Supplementives, the Coordination Account and Conflicting Intentions." *Philosophical Perspectives* (27): 288–311.  
<https://doi.org/10.1111/phpe.12028>
- Kölbl, Max. 2003. "Faultless Disagreement." *Proceedings of the Aristotelian Society*: Supplementary Volume 104: 53–73. [doi.org/10.1111/j.0066-7373.2004.00081.x](https://doi.org/10.1111/j.0066-7373.2004.00081.x)
- Kolodny, Niko. 2018. "Instrumental Reasons." In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star, 731–64. Oxford: Oxford University Press.
- Lasersohn, Peter. 2005. "Context Dependence, Disagreement, and Predicates of Personal Taste." *Linguistics and Philosophy* 28: 643–86.  
[doi.org/10.1007/s10988-005-0596-x](https://doi.org/10.1007/s10988-005-0596-x)
- MacFarlane, John. 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press.
- Marques, Teresa. 2014. "Disagreeing in Context." *Frontiers in Psychology* (6): 1–12. <https://doi.org/10.3389/fpsyg.2015.00257>
- Marques, Teresa. 2016. "We Can't Have No Satisfaction." *Filosofia Unisinos* 17 (3): 308–14. <https://doi.org/10.4013/fsu.2016.173.07>
- Marques, Teresa, and Manuel García-Carpintero. 2014. "Disagreement about Taste: Commonality Presuppositions and Coordination." *Australasian Journal of Philosophy* 92 (4): 701–23. [doi.org/10.1080/00048402.2014.922592](https://doi.org/10.1080/00048402.2014.922592)
- Michaelson, Eliot. 2014. "Shifty Characters." *Philosophical Studies* 167 (3): 519–40.  
<https://doi.org/10.1007/s11098-013-0109-7>
- Palmira, Michele. 2015. "The Semantic Significance of Faultless Disagreement." *Pacific Philosophical Quarterly* 96 (3): 349–71. [doi.org/10.1111/papq.12038](https://doi.org/10.1111/papq.12038)
- Richard, Mark. 2004. "Contextualism and Relativism." *Philosophical Studies* 119 (1-2): 215–42. [doi.org/10.1023/B:PHIL.0000029358.77417.df](https://doi.org/10.1023/B:PHIL.0000029358.77417.df)
- Schaffer, Jonathan. 2011. "Perspective in Taste Claims and Epistemic Modals." In *Epistemic Modality*, edited by Andy Egan, and Brian Weatherson, 179–226. Oxford: Oxford University Press.
- Schiffirin, Deborah. 1984. "Jewish Argument as Sociability." *Language in Society* 13 (3): 311–35. <https://doi.org/10.1017/S0047404500010526>
- Solt, Stephanie. 2018. "Multidimensionality, Subjectivity and Scales: Experimental Evidence." In *The Semantics of Gradability, Vagueness and Scale Structure*.

- Experimental Perspectives, edited by Elena Castroviejo, Louise McNally, and Galit Weidman Sassoon, 59–91. Cham, Switzerland: Springer.
- Stanley, Jason. 2005. *Knowledge and Practical Interests*. Oxford: Oxford University Press.
- Stojanovic, Isidora. 2007. “Talking about Taste: Disagreement, Implicit Arguments, and Relative Truth.” *Linguistics and Philosophy* 30 (6): 691–706. [doi.org/10.1007/s10988-008-9030-5](https://doi.org/10.1007/s10988-008-9030-5)
- Stojanovic, Isidora. 2019. “Disagreements about Taste vs. Disagreements about Moral Issues.” *American Philosophical Quarterly* 56 (1): 29–41. <https://www.jstor.org/stable/45128641>
- Sundell, Timothy. 2011. “Disagreements about Taste.” *Philosophical Studies* 155: 267–88. <https://doi.org/10.1007/s11098-010-9572-6>
- Sundell, Tim. 2016. “The Tasty, the Bold and the Beautiful.” *Inquiry* 59 (6): 793–818. <https://doi.org/10.1080/0020174X.2016.1208918>
- Umbach, Carla. forthcoming. “Evaluative Predicates beyond ‘fun’ and ‘tasty’”. In *The Wiley Blackwell Companion to Semantics*, edited by Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann, and & Thomas Ede Zimmerman. Hoboken: Wiley-Blackwell.
- Wright, Crispin. 2001. “On Being in a Quandary: Relativism, Vagueness, Logical Revisionism.” *Mind* 110: 45–98. [doi.org/10.1093/mind/110.437.45](https://doi.org/10.1093/mind/110.437.45)
- Zakkou, Julia. 2019. *Faultless Disagreement: A Defense of Contextualism in the Realm of Personal Taste*. Frankfurt: Verlag Vittorio Klostermann.

# ‘Boys Don’t Cry’ – An Ambiguous Statement?

Katharina Felka\*

Received: 16 November 2020 / Revised 27 April 2021 / Accepted: 19 May 2021

*Abstract:* As has often been observed in the literature, an utterance of a generic such as ‘Boys don’t cry’ can convey a normative behavioural rule that applies to boys, roughly: that boys shouldn’t cry. This observation has led many authors to the claim that generics are ambiguous: they allow both for a descriptive as well as a normative reading. The present paper argues against this common assumption: it argues that the observation in question should be addressed at the level of pragmatics, rather than at the level of semantics. In particular, the paper argues that the normative force of utterances of generics results from the presence of a conversational implicature. This result should somewhat alleviate the task of finding a proper semantic analysis of generics since it shows that at least one of their intriguing features need not be reflected in their truth-conditions.

*Keywords:* Generics; normative generics; semantics-pragmatics interface; conversational implicatures.


## 1. Introduction

Generics are statements such as ‘Tigers have stripes’, ‘Birds eat worms’, or ‘Houses have doors’. In a first approximation, we can say that they are

---

\* University of Graz

 <https://orcid.org/0000-0002-4921-8815>

 Institute of Philosophy Heinrichstr. 26/5th floor, A-8010 Graz, Austria

 [katharina.felka@uni-graz.at](mailto:katharina.felka@uni-graz.at)



statements that express *generalisations*, in the present examples, generalisations about tigers, birds, or houses.<sup>1</sup> But it is far from trivial to say something more precise about generics, in particular to give a precise statement of their truth-conditions. A natural thought appears to be that generics function similarly to quantified statements such as ‘All tigers have stripes’, ‘Most tigers have stripes’, or ‘Some tigers have stripes’. That is, it appears natural to assume that a *certain amount* of the members of the pertinent kind need to have the property in question for a generic to be true. But it is unclear what that amount could be. Some generics appear to require the majority of the members of the pertinent kind to have the property in question. For instance, the truth of the generics ‘Tigers have stripes’ and ‘Birds eat worms’ appears to require most tigers to be striped and most birds to eat worms respectively. But others do not. For instance, the generic ‘Ducks lay eggs’ is true even though only the mature female ducks lay eggs. This makes it questionable whether a semantic analysis of generics can be given in terms of a quantified determiner.<sup>2</sup>

A further complication in spelling out the truth-conditions of generics results from the observation that some generics seem to allow for descriptive as well as normative readings. A paradigm example is the generic ‘Boys don’t cry’. As has often been observed in the literature, this generic can convey a descriptive generalisation about boys. But it can also convey a normative behavioural rule that applies to boys, roughly: that boys *shouldn’t* cry. This observation has led many authors to the claim that generics are ambiguous. Some claim that their ambiguity results from the fact that they can exhibit different logical forms, while others have argued that generics are not *per se* ambiguous. According to them, rather nouns such as ‘boys’ are ambiguous and, thus, generics that contain such nouns are only *derivatively* ambiguous. Both accounts, however, share the assumption that the observation in question needs to be addressed at the level of *semantics* and, thus, should be reflected in the truth-conditions of generics.

---

<sup>1</sup> The examples are all bare plural generics. Generics can also contain a definite or indefinite description instead of a bare plural (e.g., ‘The tiger has stripes’ or ‘A tiger has stripes’).

<sup>2</sup> Cp. Leslie and Lerner (2016), amongst many others.

The present paper will argue against this common assumption: it will argue that the observation in question—that generics appear to allow for descriptive as well as normative readings—should be addressed at the level of *pragmatics* rather than at the level of semantics. In particular, the paper will argue that the normative force of utterances of generics does not result from ambiguity of one of the uttered expressions—neither of the generic itself nor of an expression contained in the generic. Rather, it arises due to the presence of a *conversational implicature*. If correct, the result of the paper should somewhat alleviate the task of finding a proper semantic analysis of generics, since it shows that at least one of the intriguing features of generics need not be reflected in their truth-conditions.

The structure of the paper is as follows. In section 2 I will present further examples to motivate the claim that generics allow for descriptive as well as normative readings. In section 3 I will present two semantic explanations that have been proposed in the literature and argue that they both cannot fully capture the extent of the phenomenon in question. In section 4 I will propose a pragmatic alternative and argue that it is superior to the semantic explanations.

## 2. Descriptive and normative reading

Generics are usually used to express generalisations. For instance, 'Tigers have stripes' expresses a generalisation about tigers and 'Birds eat worms' expresses a generalisation about birds. However, as already indicated, not all utterances of generics seem to express a descriptive generalisation. Some utterances of generics have 'a certain kind of normative force' (Leslie 2015, 112). Consider, for instance, the following quote from the TV show *Breaking Bad*, in which Gus Fring, a drug dealer, tries to convince Walter White, who is in urgent need of money, to continue selling drugs for him:

And a man, a man provides. And he does it even when he's not appreciated, or respected, or even loved. He simply bears up and he does it. Because he's a man.

In this quote Gus Fring uses the generic 'A man provides'. But he does not use it to make a descriptive generalisation about what a man in general does.

Rather, his utterance has normative force: it is supposed to convey that the addressee has the obligation to provide for his family.

There are many more similar examples. For instance, utterances of generics such as ‘Boys don’t cry’, ‘Women put family before career’, ‘Women don’t wear pants’, ‘Men open doors for ladies’, ‘Children at our day care don’t hit each other’, or ‘Friends don’t let friends drive drunk’ are usually not used to express generalisations about boys, women, men, children, or friends. This becomes particularly obvious from the fact that one may sincerely utter the generics even though one does not have any opinion about the actual distribution of the property in question among boys, women, men, children, or friends respectively—or even if one thinks that boys do cry, women do not put family before career, or friends do let friends drive drunk. As Leslie puts it, the generic ‘Friends don’t let friends drive drunk’ is not ‘a banal descriptive observation; utterances of it rather serve as *injunctives* precisely because friends ... all too often let their friends drive drunk ...’ (Leslie 2015, 134). Similarly, the generics ‘Women don’t wear pants’ or ‘Children at our day care don’t hit each other’ might be uttered in situations in which the speaker thinks that women do wear pants or that children at the day care do hit each other to point out that this behaviour is in conflict with a norm he takes to be in place.

While this phenomenon has been noted by many authors, there are only a few authors that have tried to provide an explanation for it.<sup>3</sup> According to the explanations that have been proposed so far, generics are in some way ambiguous. This claim gives rise to the question of what the pertinent readings of generics consist in and what exactly induces their ambiguity. In the following I will discuss two accounts that answer these questions, the first given by Sarah-Jane Leslie and the second given by Ariel Cohen.

---

<sup>3</sup> Cp., e.g., Burton-Roberts (1977), Carlson (1995), Cohen (2001), Greenberg (2003), McConnell-Ginet (2012), Krifka (2013), and Leslie (2015). While Leslie (2015) is interested in the question what induces the normative force of a generic like ‘Boys don’t cry’, the other authors are primarily concerned with a puzzle about the distribution of indefinite singular and bare plural generics. However, some of the proposed solutions (in particular Cohen 2001, Greenberg 2003, and Krifka 2012) can be extended to answer the question at hand and are thus discussed in the following as well.



### 3. Semantic explanations

Both Leslie's and Cohen's accounts rely on the assumption that generics are ambiguous but they differ in how they explain their ambiguity. While Leslie has argued that their ambiguity results from the nouns they contain, Cohen has claimed that it rather results from the logical structure of generics themselves. In the following I will critically discuss both accounts, starting with Leslie's.

#### 3.1 *Social role nouns*

Following work by Knobe et al. (2013), Leslie (2015) has argued that nouns such as 'boys', 'women', 'men'—in general: nouns that denote groups of people for which certain social norms are in place (in the following: *social role nouns*)—are ambiguous. For instance, the noun 'boys' can either be used (roughly) in the sense of 'premature human beings that have the biological characteristics of males' or (roughly) in the sense of 'human beings who fulfil the ideals of boyhood'. Accordingly, a generic such as 'Boys don't cry' can be used in two senses as well: it can either be used in the sense of 'Premature human beings that have the biological characteristics of males don't cry' or in the sense of 'Human beings that fulfil the ideals of boyhood don't cry' (or shortly 'Ideal boys don't cry').

There are two difficulties with Leslie's account. Firstly, it is questionable whether the claim that social role nouns are ambiguous is sufficient in order to explain the normative force that generic utterances can have. According to Leslie, a normative utterance of 'Boys don't cry' expresses the content that ideal boys don't cry. Clearly, an utterance with this content *can* have normative force, e.g. in a context in which it is part of the common ground that the addressee wants to be an ideal boy. In such a context, the addressee can simply infer that he should not cry in order to count as an ideal boy. However, an utterance of 'Boys don't cry' can have normative force even in a context in which this proposition is not part of the common ground, e.g. in a context in which it is commonly assumed that the addressee does not want to be a boy, let alone be an ideal boy. In such a context it is unclear how conveying a generalisation about ideal boys can serve as an injunction

that has normative force. Hence, as it stands, Leslie's account cannot fully accommodate the phenomenon in question.<sup>4</sup>

Secondly, Leslie's account predicts that only generics that contain social role nouns can be used in two different ways. But that prediction is not borne out. This can already be seen from some of the examples that Cohen (2001, 194) presents. Among others, he cites the generic 'Bishops move diagonally'. As Cohen points out, this generic gives rise to the same phenomenon as generics like 'Boys don't cry'. But the noun 'bishops' is not a social role noun: it does not denote a group of people for which certain social norms apply. Accordingly, the pertinent normative reading cannot be due to the fact that the generic contains an ambiguous social role noun. Further examples that illustrate the same point are easy to find. Suppose, for instance, a young interior designer proposes a sparse and clean interior for a new house and his superior rejects his proposal with the words 'Family homes are warm and cosy'. Surely, the point of the superior's statement is not to inform her employee about how family homes are generally designed—she may not even believe that in general family homes are warm and cosy. Rather, she informs her employee how he *should* make family homes look like. But, again, 'family home' is not a social role noun and, thus, the normative force of the utterance cannot be due to the fact that it contains an ambiguous social role noun.

In order to deal with this difficulty, Leslie might extend her account to nouns like 'bishops' or 'family homes', i.e. she might claim that such nouns are ambiguous as well. But, firstly, this would require her to postulate even more widespread ambiguities since presumably *any* generic can receive a normative reading in appropriate circumstances. Secondly, extending her account in this way would still not allow her to account for the extent of the phenomenon. Take, for instance, the quantified sentence 'None of the other boys is crying.' In uttering this sentence one can simply describe what the other boys are doing. But one can also call on someone to behave in

---

<sup>4</sup> Leslie could appeal to further linguistic mechanisms. For instance, she could argue that in such a context the normative force of the utterance is due to a Gricean implicature. But the following discussion will make clear (i) that Leslie would object to such an extension of her account and (ii) that there is a more parsimonious account available.

a certain way: if a mother says to her crying son ‘None of the other boys is crying’, then she conveys that her son *should* behave like the other boys do.<sup>5</sup> But in this case the normative force of the utterance cannot be due to the fact that the mother uses ‘boys’ in the sense of ‘ideal boys’ since she does not express that the other ideal boys don’t cry. Leslie might try to argue that the linguistic mechanism that is at work in this case differs from the one that is pertinent in the case of generics. But it is not clear that this challenge can be met: in both cases we have a sentence that can either be used to describe something or that can serve as an injunction with normative force—and in both cases the normative force cannot be due to the fact that the sentence contains any obviously normative vocabulary. Without any further indication, it is thus not clear why the cases should be kept apart.

To sum up, Leslie’s account is insufficient since (i) it cannot fully explain the normative force that generic utterances can have and (ii) it fails to account for the extent of the phenomenon: generics that do not contain social role nouns allow for normative as well as descriptive readings as well. Of course, it still seems to be the case that generics that contain social role nouns are especially susceptible for allowing a descriptive as well as a normative reading. A complete account of the phenomenon should accommodate this observation. In due course, I will try to provide such an account.

### 3.2 *Divergent logical forms*

According to Cohen (2001), generics themselves are ambiguous: they can have different logical forms. For instance, the generic ‘Boys don’t cry’ can express a generalisation and, thus, have (roughly) the following logical form:<sup>6</sup>

$$\text{Gen } x [\text{Boys } (x)][\text{Do not cry } (x)]$$

In this case the operator ‘Gen’ functions as a generic operator for which Cohen proposes a quantificational analysis. However, the generic ‘Boys don’t cry’ can also have (roughly) the following logical form:

---

<sup>5</sup> Sterken (2014, 162) also notes that quantificational sentences can have a ‘rules-and-regulations’ reading as well.

<sup>6</sup> I follow here Leslie’s simplified presentation of Cohen’s account (cp. Leslie 2015, 119).

*in – effect!* ( $Boys(x) \Rightarrow Do\ not\ cry(x)$ )

In this case the generic ‘Boys don’t cry’ ascribes to a certain rule the property of being in effect, namely to the rule that boys don’t cry. Since a generic like ‘Boys don’t cry’ can have either of these two logical forms, it can be used to express a generalisation or it can be used to express that a certain rule is in effect. Or so Cohen claims.

In contrast to Leslie’s account, Cohen’s account has the advantage that it can deal with generics like ‘Bishops move diagonally’, which do not contain a social role noun. According to his account, *any* generic can either be used to express a generalisation or to express that a certain rule is in effect. And, yet, it appears that his account is confronted with a similar difficulty as Leslie’s account: it cannot capture the extent of the phenomenon in question. For there appear to be *non-generic* statements that can be used either descriptively or with ‘a certain kind of normative force’ (Leslie 2015, 112). As already pointed out, this holds for quantified statements like ‘None of the other boys is crying’. The same observation can be made with respect to singular statements: suppose, for instance, it is a general rule in a farmer’s family that the first-born son takes over the farm when he is old enough. Let us further suppose that the first-born son would prefer to leave the countryside and his sister, who endorses his plan, tries to put in a good word for him. But when she tries to convince her parents, her father only replies: ‘He takes care of the farm’. In this case, the father conveys his opinion about what the son is supposed to do—and he does so by uttering a sentence that is usually used to *describe* what someone does. Hence, non-generic statements give rise to exactly the same phenomenon: they can be used to describe something but they can also serve as an injunction even though they do not contain any obviously normative vocabulary.

Thus, both Leslie and Cohen overlook that the phenomenon in question can arise for almost any kind of statement. However, this observation suggests that a pragmatic explanation is preferable to a semantic one that ties it to some special kind of words or construction. In the following section I will propose such a pragmatic explanation and defend it against objections.<sup>7</sup>

---

<sup>7</sup> Krifka (2012) has proposed a semantic explanation as well. According to Krifka, the normative force of some utterances of generics is due to the fact that generics

#### 4. A pragmatic explanation

As a general methodological rule, Grice (1989, 47) has recommended not to multiply meanings beyond necessity. Following his advice, we should investigate whether we can give a pragmatic explanation of the phenomenon in question. According to such an explanation, utterances of generics such as 'Boys don't cry' express a generalisation about boys and in some contexts they additionally conversationally implicate that in a certain respect (crying) the addressee *should* behave like boys in general do. This explanation is admittedly quite simple but it has the following advantages in contrast to Leslie's and (partly) Cohen's explanations:

- (1) *Normative force* A pragmatic explanation does not have any trouble to explain that an utterance of 'Boys don't cry' can have normative force even in utterance contexts in which it is part of the common ground that the addressee does not want to be a boy, let alone be an ideal boy. For, according to this explanation, the speaker implicates that the addressee should behave like boys in general do—and that implicature can have normative force even if the addressee does not want to be a boy.
- (2) *Parsimony* A pragmatic explanation does not postulate additional meanings and, thus, keeps the lexicon simple.
- (3) *Generality* According to a pragmatic explanation, the phenomenon in question results from features of utterance contexts and, thus, it can account for the fact that it is not restricted to a *certain kind of expression* (social role nouns) or to a *certain kind of construction* (generics).

---

allow for a definitional reading: e.g. a normative utterance of 'Boys don't cry' expresses that the concept *boys* is defined such that only non-crying things fall under that concept (or at least that it should be defined in such a way; cp. Greenberg (2003) for a similar account). But again: (i) this claim alone is not sufficient to explain the normative force of such utterances since there are contexts in which mere information about the definition of *boys* is not suitable to direct the addressees' actions; (ii) this claim does not appear to be able to capture all normative uses of generics. For instance, a speaker's utterance of 'A tiger lives outdoors' can have normative force but in making that utterance the speaker certainly does not want to claim that living outdoors is part of the definition of the concept *tiger*.

And yet it allows us to explain why the phenomenon occurs more often with generics that contain social role nouns rather than with other kinds of expressions. Firstly, it occurs more often with generics (rather than with singular statements such as ‘He works at the farm’) since generics say something about groups of people (or things) and norms and rules are usually formulated for *groups* of people (or things), rather than for singular entities. And, secondly, it occurs more often with generics that contain social role nouns rather than with generics that do not (‘Family homes are warm and cosy’) since most norms and rules that are in place apply to *social* groups, rather than to groups of things like homes or animals.

Thus, *prima facie* it appears that a pragmatic explanation is superior to extant alternatives. However, Leslie (2015) has provided two objections against a pragmatic explanation. In the following I will discuss her objections as well as a further worry a pragmatic explanation might give rise to.

#### 4.1 *How the implicature is triggered*

It is generally assumed that if something is a conversational implicature, its presence has to be explicable by appeal to conversational maxims. However, according to Leslie, this requirement is not fulfilled:

Consider, for example, a standard utterance of ‘friends don’t let friends drive drunk’. For the pragmatic account to explain its normative force, we would have to suppose that [it] is so obviously false as a descriptive statement that the speaker could not have possibly meant to assert that—or alternatively so obviously true that it triggers a search for a more informative content. Neither characterization seems remotely plausible. (Leslie 2015, 137)

Leslie argues that neither the maxim of quality nor the maxim of quantity can explain the presence of an implicature. While this appears to be correct, there is a further maxim—the maxim of relation (‘Be relevant’)—that is perfectly well able to explain the presence of a conversational implicature. Suppose, for instance, a mother says to her crying son ‘Boys don’t cry’. If in making this utterance the mother were only conveying a descriptive generalisation about what boys in general do, her utterance would be

irrelevant in the utterance context. Accordingly, the boy will try to interpret her statement in a way such that it is relevant and thereby arguably come to the conclusion that she wants to convey that in a certain respect (crying) he should behave like boys in general do. Thus, the presence of the implicature *can* be explicated by appeal to one of the conversational maxims, it is simply a different maxim than the ones Leslie considers.

4.2 *Empirical counter evidence?*

Leslie’s second argument relies on the assumption that children are able to grasp the normative force that an utterance of ‘Boys don’t cry’ may have. However, so she proceeds, children are not able to understand conversational implicatures. Thus, she concludes, the normative force that such an utterance may have cannot be due to the presence of an implicature.

Leslie cites an experiment due to Noveck (2001) to justify the premise that children are not able to understand conversational implicatures. In this experiment 60 children and 25 adults were asked to either accept or reject statements containing the determiner ‘some’. Some of the statements were false (‘Some stores are made of bubbles’), some true (‘Some birds live in cages’), and some true but infelicitous (‘Some giraffes have long necks’). As you can see in the table, in contrast to the adults, most children accepted the statement ‘Some giraffes have long necks’ even though it conflicts with the maxim of quantity (‘Be informative’).<sup>8</sup> Based on this experiment, Leslie claims that there is strong empirical evidence that children do not understand conversational implicatures.

Statements	Evaluation	8 years	10 years	Adults
False (‘Some stores are made of bubbles’)	reject	95%	99%	98%
True and felicitous (‘Some birds live in cages’)	accept	84%	90%	99%
True but infelicitous (‘Some giraffes have long necks’)	accept	89%	85%	41%

<sup>8</sup> This utterance conflicts with the maxim of quantity since a more informative statement could have been made: ‘All (or at least most) giraffes have long necks’.

However, it is doubtful whether Noveck's experiment indeed provides support for Leslie's premise. Firstly, if at all, it could only show that children are not able to understand scalar implicatures, i.e. implicatures that arise due to the *maxim of quantity*. But if the pragmatic explanation proposed above is correct, they only need to be able to understand implicatures that arise due to the *maxim of relation*. Secondly, Noveck's findings are questionable. Chierchia et al. (2000) conducted a similar experiment and came to a different conclusion. In their experiment they told 15 children and 8 adults a story in which four boys had to choose between a skateboard and a bicycle. After telling the story, a puppet made the true and felicitous statement 'Each of the four boys chose a skateboard and a bicycle' while another puppet made the true but infelicitous statement 'Each of the four boys chose a skateboard or a bicycle'. The participants in the experiment had then the task to say which puppet described better what happened in the story. In this experiment there were no significant deviances: both the adults and the children said in almost all cases that the first puppet said it better. Based on these findings, Chierchia et al. conclude that children *do* understand scalar implicatures. According to their hypothesis, the deviances in Noveck's experiment are rather due to the fact that children cannot keep in mind formulations long enough in order to compare them with alternative formulations (i.e. when they only hear 'Some giraffes have long necks', they are not aware of the fact that one may also make the more informative statement 'All (or at least most) giraffes have long necks').

Thus, Leslie's second objection is not convincing either: firstly, it is doubtful whether the given empirical evidence is pertinent at all and, secondly, it is not robust enough in order to support the central premise of her objection.

#### 4.3 A further worry

According to the explanation presented above, a (normative) utterance of the generic 'Boys don't cry' expresses that in general boys don't cry and additionally implicates that the addressee should behave like boys in general do via the maxim of relation. This explanation might give rise to a worry: conversational implicatures that arise due to the maxim of relation are usually *additive implicatures*, rather than *substitutional implicatures*. Additive



implicatures are ones that are conveyed *in addition* to the semantic content expressed, while substitutional implicatures are conveyed *instead* of the semantic content expressed (paradigm examples are, e.g., ironic statements that convey the opposite of the expressed content due to the maxim of quality).<sup>9</sup> However, as pointed out at the outset, one may use a generic like 'Boys don't cry' with normative force without believing and, thus, without expressing that in general boys don't cry. In such cases the implicature in question would have to be a substitutional implicature. But it is questionable whether there are any substitutional implicatures that arise due to the maxim of relation. Or so the worry goes.

However, the account presented here is not committed to the claim that in the cases at hand the normative force of generics is due to the presence of a substitutional implicature. For it appears natural to assume that just like *any other* kind of statement containing nouns (e.g. 'I want you to be a boy') the noun in generics like 'Boys don't cry' can be accompanied by the (pronounced or unpronounced) modifier 'real' (or 'ideal') that allows us to speak of only a subgroup of the group denoted by the noun. If a speaker sincerely utters 'Boys don't cry' even though she does not share the view that in general boys don't cry, then her utterance presumably contains the unpronounced modifier 'real' and, hence, expresses the content that real boys don't cry.<sup>10</sup> If so, the implicature in question—that the addressee should behave like a real boy—is an additive implicature just like in the other cases.

Thus, in order to fully account for the phenomenon at hand we need to assume that in some cases nouns like 'boys' are accompanied by the

---

<sup>9</sup> Cp. Meibauer (2009) for the distinction between additive and substitutional implicatures, among others.

<sup>10</sup> According to Leslie, the phrase 'real Fs' cannot have the function to speak of a subgroup of the Fs. For, Leslie says, we can correctly use it in a statement like 'Hilary Clinton is the only [real] man in the Obama administration' even though Hilary Clinton is a woman (cp. Leslie 2015, 115). However, this observation can be explained in pragmatic terms again: in making this utterance one expresses an obvious falsehood and, thus, the listener will search for some other content that is conveyed (in the present case: that Clinton is the only person in the Obama administration that has the pertinent features of real men). Thus, Leslie's observation does not present any reason to depart from the standard view, according to which 'real' functions as a modifier.

pronounced or unpronounced modifier ‘real’. But this assumption is harmless since it is an assumption we are readily willing to make with respect to other kinds of statements as well. Further, the resulting account is still not confronted with the difficulties that arise for Leslie’s and Cohen’s account. Firstly, in contrast to Leslie’s account, it does not conflict in any way with Ockham’s razor since the modifier ‘real’ is already contained in our lexicon and we thus do not have to add any further entries to our lexicon. Secondly, in contrast to Leslie’s and Cohen’s account, the account can accommodate the extent of the phenomenon since it does not rely on the assumption that the normative force of an utterance of ‘Boys don’t cry’ is due to some specific kind of words contained in the utterance or to the construction itself. According to the account presented here, its normative force lies entirely in a Gricean implicature (i.e., that one should behave like a (real) boy) and, thus, arises at the level of pragmatics.

## 5. Conclusion

As has often been observed in the literature, generics such as ‘Boys don’t cry’ allow for a descriptive as well as for a normative reading. According to the explanations that have been provided so far, this observation should be accounted for at the level of semantics. In contrast, the present paper has argued that we should rather account for it in terms of conversational implicatures that arise due to the maxim of relation. This proposal somewhat lightens the difficult task of finding the proper semantics of generics since it shows that at least one observation regarding generics need not be accounted for at the semantic level.

## References

- Burton-Roberts, Noel. 1977. “Generic Sentences and Analyticity.” *Studies in Language* 1 (2): 155–96. <https://doi.org/10.1075/sl.1.2.02bur>
- Carlson, Gregory N. 1995. “Truth-Conditions of Generic Sentences: Two Contrasting Views.” In *The Generic Book*, edited by Gregory N. Carlson, and Francis Jeffrey Pelletier, 224–37. Chicago: Chicago University Press.
- Chierchia, Gennaro, Stephen Crain, Maria Teresa Guasti, Andrea Gualmini, and Luisa Meroni. 2001. “The Acquisition of Disjunction: Evidence for

- a Grammatical View of Scalar Implicatures." In *Proceedings of the 25<sup>th</sup> Boston University Child Language Development Conference*, edited by Anna H.-J. Do, Laura Domínguez, and Aimee Johansen, 157–68. Cascadilla Press.
- Cohen, Ariel. 2001. "On the Generic Use of Indefinite Singulars." *Journal of Semantics* 18 (3): 183–209. <https://doi.org/10.1093/jos/18.3.183>
- Greenberg, Yael. 2003. *Manifestations of Genericity*. New York: Routledge.
- Grice, Paul. 1989. *Studies in the Way of Words*. Cambridge MA: Harvard University Press.
- Knobe, Joshua, Sandeep Prasada, and George E. Newman. 2013. "Dual Character Concepts and the Normative Dimension of Conceptual Representation." *Cognition* 127 (2): 242–57. <https://doi.org/10.1016/j.cognition.2013.01.005>
- Krifka, Manfred. 2013. "Definitional Generics." In *Genericity*, edited by Alda Mari, Claire Beyssade, and Fabio Del Prete, 372–89. Oxford: Oxford University Press.
- Leslie, Sarah-Jane. 2015. "'Hillary Clinton is the Only Man in the Obama Administration': Dual Character Concepts, Generics, and Gender." *Analytic Philosophy* 56 (2): 111–41. <https://doi.org/10.1111/phib.12063>
- Leslie, Sarah-Jane, and Adam Lerner. 2016. "Generic Generalizations." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. <https://plato.stanford.edu/entries/generics/>
- McConnell-Ginet, Sally. 2012. "Generic Predicates and Interest-Relativity." *Canadian Journal of Linguistics* 57: 261–87. <https://doi.org/10.1017/S0008413100004771>
- Meibauer, Jörg. 2009. "Implicature". In *Concise Encyclopedia of Pragmatics* (second edition), edited by Jacob L. Mey, 365–78. Amsterdam: Elsevier.
- Noveck, Ira. 2001. "When Children are More Logical than Adults: Experimental Investigations of Scalar Implicatures." *Cognition* 78 (2): 165–88. [https://doi.org/10.1016/S0010-0277\(00\)00114-1](https://doi.org/10.1016/S0010-0277(00)00114-1)
- Sterken, Rachel. 2014. *Generics in context: Generalisation, Context and Communication*. Ph.D. thesis, University.

## Value and Scale: Some Observations and a Proposal

Andrés Soria-Ruiz\*

Received: 3 December 2020 / Revised: 7 May 2021 / Accepted: 12 June 2021

*Abstract:* In this paper, I investigate the scalar semantics of evaluative adjective in general, and of *good* in particular. Lassiter (2017) has argued that *good*, when taking propositions as arguments, has an interval scale. I argue that there's evidence in support of the view that *good*, when taking individuals as argument, has a scale that is stronger than interval, but weaker than ratio. In particular, I propose that individual-level *good* has a “round” ratio scale, which allows a broader set of ratio transformations than standard ratio scales. This conclusion is consistent with the fact that *good* admits round ratio modifiers (*twice as good*), but eschews precise ones (*# 1.38x as good*). An important consequence of this view is that the scales of individual and propositional-level *good* are severed.

*Keywords:* Evaluative adjectives, scalar semantics; metaethics.

“All my life I'd heard people tell their black boys and black girls to “be twice as good,” which is to say “accept half as much.” These words would be spoken with a veneer of religious nobility, as though they evidenced some unspoken quality, some undetected courage, when in fact all they evidenced was the gun

---

\* Universidade NOVA de Lisboa

 <https://orcid.org/0000-0002-4592-9783>

 Campus de Campolide 1099-085 Lisboa, Portugal.

 [aruiz@fcs.unl.pt](mailto:aruiz@fcs.unl.pt)

to our head and the hand in our pocket.” (Ta-Nehisi Coates, *Between the World and Me*)

## 1. Introduction

This paper puts forward a puzzle about the semantics of evaluative adjectives, in particular about the adjective *good*. The puzzle is the following: even though *good* largely eschews measurement, phrases like *twice as good* are perfectly interpretable. What do they mean? And what consequences does their acceptability have for the semantics of these adjectives? The purpose of this paper is to investigate those questions.

Evaluatives<sup>1</sup> are gradable predicates, which is attested by the fact that they admit ADJECTIVAL MODIFIERS. To see this, compare (1), where a gradable adjective (*good*) is modified, with (2), where a non-gradable adjective (*hexagonal*) is modified (# indicates that the construction is unacceptable):

- (1) The courtyard is {very} good / {much / a little / better than the park}.
- (2) The courtyard is {# very} hexagonal / {# much / # a little / # more hexagonal than the park}.<sup>2</sup>

Adjectival modification is a window into the scalar properties of gradable adjectives (Lassiter 2017; Sassoon 2010; Solt 2018, a.m.o.). Different modifiers can tell us different things about the scale corresponding to the relevant adjective. For instance, an adjective like *tall* admits measure phrases, while *good* does not:

- (3) Ann is 180cm tall.
- (4) Bill is ??? good.

<sup>1</sup> This is a heterogeneous class of adjectives whose most eminent members are *good* and *bad*, but which also contains adjectives of moral (*virtuous*), aesthetic (*beautiful*) and personal taste evaluation (*tasty*), as well as so-called thick adjectives (*cruel*).

<sup>2</sup> This does not mean that the sentences in (2) are absolutely unintelligible; but in order to recover a meaning one needs to do some interpretative work. For instance, *the courtyard is a lot more hexagonal than the park* could mean that it has a more regular and/or carefully delineated hexagonal shape.

Interestingly, one would not even know what to fill the blank with. Bill is... “2 hours good”? “4 agreeable encounters good”? This suggests that, while the scale corresponding to *tall*, i.e., height, admits of measurement, the scale corresponding to *good* does not. There are no standard measures of how good things are.<sup>3</sup>

Regardless, I want to argue that *good* does not fully eschew measurement. Indeed, my purpose is to show that the scale of *good* poses a puzzle. Among the two most salient scale types used in social sciences, ratio and interval, it is difficult to determine which of these corresponds to *good*. On the one hand, if *good* had an interval scale, it ought to reject ratio modifiers. But individual-level *good* (*x is good*) admits round ratio modifiers (viz. Coates’ quote). On the other hand, if *good* had a ratio scale, it ought to be positive with respect to concatenation. This means, roughly, that the combined goodness of any two individuals taken together must be greater than the goodness of each individual taken separately. But according to Lassiter (2017), propositional-level *good* (*it is good that φ*) is not positive with respect to concatenation. Thus, one is confronted with a puzzle. The way out will be to assume that *good* has a different scale when it takes propositions

---

<sup>3</sup> There are exceptions: one can speak of a swimmer being 6 seconds better than another; or a politician doing 3 points better than their opponent on a poll. However, it is intuitive to interpret *better* in those examples as meaning 6 seconds, or 3 points greater on some contextually salient scale, which may not be the scale of *good*. The presence of specific units of measurement in those examples indicates as much: *6 seconds better* suggests that better there just means *faster*; *3 points better* indicates that *better* stands for *greater*. Such reinterpretations can coerce *good* into admitting exact measurements, and moreover, into shifting its scale-type: in both examples, *better* will adopt a ratio scale in virtue of the fact that the relevant properties (speed, score on a poll) have ratio scales. Examples abound: *Apple performed ... 4.5 times better than Blackberry* (<https://www.tradegecko.com/blog/supply-chain-management/apple-the-best-supplychain-in-the-world>). Relatedly, Lassiter (2017, p.89) discusses an Internet example where a company is described as retaining users *2-3 times as efficiently* as another. Independently of whether *efficient* or *good* really are ratio adjectives, in these contexts they behave as such and thereby admit the relevant ratio modifiers. I will not, however, rely on examples like these to conclude that *good* allows exact measurement or has a ratio scale. I thank two reviewers for pressing me to clarify this. In what follows, I set such coerced usages aside.

and when it takes individuals as arguments. In the former case, I will conclude with Lassiter that *good* has an interval scale. In the latter case, however, there are reasons to conclude that *good* doesn't have an interval scale. More specifically, I will propose that individual-level *good* has a scale that is stronger than interval, but weaker than ratio, a scale which I call ROUND RATIO scale, which admits a broader set of ratio transformations than standard ratio scales.

Before moving on, here is a comment on *good*. The hypothesis that *good* has a different scale-type when ranging over individuals and when ranging over propositions is bound to strike as controversial. But this may seem less surprising in light of the fact that the meaning of *good* is massively underspecified: *good* can be interpreted as categorical (*unconditional good*) or hypothetical (*good given certain ends or purposes*); relatedly, *good* has so-called "attributive" and "predicative" uses.<sup>4</sup> Moreover, *good* is judge-dependent (Bylinina 2017) and multidimensional (Sassoon 2013). Similarly, *good* carries a beneficiary argument – as *in good for you!* (see Stojanovic 2016, pp. 19-20), and even lives a double a life as an intensifier, as in *a good dose of luck* (Castroviejo and Gehrke 2019). Indeed, Hare (1952, see also Umbach 2016) held that the only thing that tied together all uses of *good* was the expression of commendation. In light of such underspecificity, the prospect of assigning different scale types to different uses of *good* may seem less striking.<sup>5</sup>

The paper is laid out as follows: in section 2, the typology of scales standardly used in linguistics is discussed, and the significance of various types of inferences and modifiers is introduced. In sections 3-5, various scale

---

<sup>4</sup> The distinction comes from Geach 1956. See e.g., Asher 2011; Ridge 2014; Thomson 2008 for discussion.

<sup>5</sup> However, this invites a further question. Given that its meaning is so open and minimal, why the focus on *good*? Other adjectives (*beautiful*, *ugly*, *interesting*) may be a bit more uniform, and thus a more reliable guide to the scalar properties of evaluative expressions. Nevertheless, there are reasons to study the semantics of *good*; after all, *good/bad* are the most basic evaluative adjectives. This is evidenced by the fact that all other evaluative adjectives imply *good/bad* in some way on another. Moreover, *good/bad* are some of the few evaluative adjectives to take both individuals and propositions as arguments, which is crucial for my discussion. I thank a reviewer for this journal for pressing me in this regard.

types are introduced (in order of increasing strength) and rejected as candidates for the scale of *good*. In section 3 it is shown that *good* cannot have a merely ordinal scale. Subsequently, I argue that the hypotheses that *good* has an interval (section 4) or ratio scale (section 5) are problematic. In section 6, I propose a solution. Section 7 concludes.

## 2. Scale structure

Lassiter (2017), Sassoon (2010), and others have resorted to Representational Measurement Theory (RMT, see Krantz et al. 1971) to explore the features of linguistically gradable items. Lassiter in particular focuses on epistemic, probability and deontic modals, but also on the evaluative adjective *good*, while Sassoon 2010 considers a more traditional set of gradable adjectives. In this and the following sections, I present the standard typology of scales following mainly Lassiter, as well as the relevant linguistic tests that can help diagnose the scale type of a scalar item, and we will see what best applies to *good*.

In RMT, the properties of scales are studied by considering what mathematical operations they support. The outcome of this is a typology of scales, or a set of scale types. Lassiter proceeds by attempting to subsume the scales lexicalized in various natural language expressions under scale types defined by RMT. His procedure is roughly the following: starting from the observation that some predicates are gradable, he assumes that they denote scalar properties, or SCALES for short. Then, in order to study the properties of those scales, he does two things. The first is to observe what kind of inferences and modifiers those natural language items allow and forbid. The second is to map the various acceptable uses of those scalar items onto different mathematical relations over the real numbers, in the way that RMT tells us to. Depending on the kind of mapping from natural language onto such mathematical relations that are admissible, a scale can be subsumed under one or other scale type.

For concreteness, let us define a SCALE as a tuple  $\mathcal{S} = \langle X, \succcurlyeq, \dots \rangle$  containing a set of individuals  $X$ , a binary ordering relation  $\succcurlyeq$  and potentially other operations. In order to determine the features of  $\mathcal{S}$ , one seeks to define a structure-preserving mapping (a homomorphism)  $\mu$  from  $\mathcal{S}$  onto  $\langle \mathbb{R}, \geq, \dots \rangle$



(where  $\mathbb{R}$  is the set of real numbers,  $\geq$  is the usual ordering relation and other operations over  $\mathbb{R}$  might be taken into account). If a function  $\mu$  is a homomorphism from  $\mathcal{S}$  onto  $\langle \mathbb{R}, \geq, \dots \rangle$ , then  $\mu$  is called an ADMISSIBLE MEASURE FUNCTION of  $\mathcal{S}$ . And to prove that  $\mu$  is an admissible measure function of  $\mathcal{S}$  is to prove a REPRESENTATION THEOREM. Different scale types are then distinguished by imposing different representation theorems that the admissible measure functions must satisfy; the more conditions they must meet, the more structure the scale has – the stronger the scale is.

There is a potentially infinite number of relations that one can define over a scale  $\mathcal{S}$ . But the crucial ones for our purposes are the binary ordering relation  $\succsim$ , which was already mentioned, and the operation of CONCATENATION, (which is represented as  $\circ$ ). Concatenation allows us to construct compound objects from the simple elements in a given domain. For any elements  $a, b$  in some domain,  $a \circ b$  is the concatenation of  $a$  and  $b$ .

However, concatenation is not a linguistic operation. In order to represent concatenation in natural language, it has to be mapped onto some model-theoretical relation. Lassiter (2017, p. 39), following Krifka (1989), maps it to the set-theoretical operation of JOIN,  $\sqcup$ , restricted to non-overlapping individuals:  $x \circ y$  is defined if and only if

1.  $x$  and  $y$  belong to the same semantic type  $\alpha$ , and
2.  $x$  and  $y$  do not overlap.

When defined,  $x \circ y = x \sqcup y$ , where  $\sqcup$  is JOIN over domain  $D_\alpha$ .

JOIN maps onto different aggregation operations depending on what domain one considers. If one considers individuals, JOIN is mereological sum. Thus, for any non-overlapping individuals  $x, y$ ,  $x \sqcup y =$  the complex individual formed by  $x$  and  $y$ ,  $x \oplus y$ . If one considers propositions, given that the JOIN operation over the domain of propositions amounts to set union, the concatenation of propositions will be their union: for any non-overlapping propositions  $u, v$ ,  $u \sqcup v = u \cup v$ , which is represented linguistically as Boolean disjunction. Concatenation is crucial because mapping it to different mathematical relations helps define different scale types (in particular ratio and interval scales).

Lassiter focuses on the three main type of scales used in social and empirical sciences, namely ORDINAL, INTERVAL and RATIO scales. These scales

are defined, as I said above, via their admissible measure functions. In order to investigate what scale a given lexical item has, one needs to consider adjectival modifiers. Adjectival modifiers carry information about the scales of the adjectives that they modify. Among the adjectival modifiers that can offer insight into scale structure there are measure phrases (*two meters, 35 years*), quantity adverbs (*much, a little, a lot*) and ratio modifiers (*twice, 1.38x*).

I will focus on patterns of acceptability and unacceptability. If an adjective accepts a modifier, then I will conclude that the adjective has a scale at least as strong as to represent the information contributed by the modifier. For example, in the introduction it was shown that (3) is an acceptable thing to say. This suggests that the scale of *tall* accepts units of measure based on centimeters (which in turn are based on ratio transformations, cf. Sassoon 2010).

On the other hand, if a modifier is not acceptable, this implies that the scalar information introduced by the modifier is too strong for the relevant adjective. For example, ratio modifiers such as *n-x* (*n-times*) require an adjective with a ratio scale to be interpretable. If they combine with an adjective that has a weaker scale, i.e., *hot, late* or *safe*, it leads to infelicity:

- (5) a. # Bowl A is 1.38x as hot as bowl B.  
 b. # Amir came 2x as late as Mora.  
 c. # My neighborhood is 4x as safe as yours.

As we will see, *hot* has an interval scale, according to which zero points are variable. *1.38x* requires a fixed zero point, and this is why it cannot combine with *hot*. The scale of *hot* does not provide a fixed zero point that *1.38x* can be interpreted relative to. Nonetheless, making explicit reference to a particular scale (e.g., Celsius) is a repair strategy for sentences like (5a):

- (6) ✓ Bowl A is 1.38x as hot as bowl B on the Celsius scale.

The reason why this qualification is successful is that making reference to the Celsius scale introduces the zero point needed to interpret the ratio 1.38.<sup>6</sup>

---

<sup>6</sup> To some speakers, sentences like (5a) sound fine. Erich Rast (p.c.) suggests the following example: *I baked the Beef Wellington twice as hot as Gordon Ramsay said:*

I turn now to presenting these three scale types and to consider whether the scale lexicalized by *good* satisfies each of them.

### 3. Ordinal scales

A scale  $\mathcal{S}$  that is merely ordinal is such that all that can be said of the elements in its domain is how they are ordered with respect to each other. For this reason, all measure functions  $\mu$  that preserve the ordering among the reals are admissible measure function of  $\mathcal{S}$ . No other structure is represented; we do not know anything about the distances between elements on the scale, for instance, or their respective distances to a zero point. The relation of set inclusion is an example of a relation with a merely ordinal structure: all the information that set inclusion represents is an ordering on its domain. More precisely:

**Definition 1** (Ordinal scale). *If a scalar property  $\mathcal{S} = \langle X, \succcurlyeq \rangle$  is an ORDINAL SCALE (disregarding concatenation and other operations), then every admissible measure function  $\mu$  that maps  $\mathcal{S}$  onto  $\langle \mathbb{R}, \geq \rangle$  is such that, for all  $x, y \in X$  and  $x \succcurlyeq y$ ,  $\mu(x) \geq \mu(y)$ .*

Any measure function  $\mu$  is an admissible measure function of  $\mathcal{S}$  as long as, to any two elements  $x, y$  of  $X$  that stand in the  $\succcurlyeq$  relation of  $\mathcal{S}$ ,  $\mu$  assigns numerical values such that the value of  $x$  is a number at least as great as  $y$ . **Definition 1** says nothing about the type of mathematical operation that concatenation should be mapped onto. Thus, any mathematical operation is admissible; it could be addition, subtraction or what have you.

To see how this works, consider again the set inclusion relation. Let us represent it as a structure  $\langle \mathcal{P}(X), \supseteq \rangle$ , where  $\mathcal{P}(X)$  is the power set of some set  $X$ , and  $\supseteq$  is the superset relation. If this structure is ordinal, then every measure function that maps it onto  $\langle \mathbb{R}, \geq \rangle$  should satisfy the representation theorem above. This implies that any mapping that respects the ordering among reals will be an admissible representation of the superset relation.

---

*I set the oven to 200 C instead of 100 C, it's burnt!* This deserves more attention, but it's possible that, in many contexts, certain scales are so prevalent that repair strategies such as (6) are not needed.

For any two elements  $Y$  and  $Z$  of  $\mathcal{P}(X)$  such that  $Y \supseteq Z$ , a  $\mu$  such that  $\mu(Y) = 5$  and  $\mu(Z) = 3$  is an admissible measure function; another  $\mu'$  such that  $\mu'(Y) = 12,351$  and  $\mu(Z) = -0.0004$  also is; but a  $\mu''$  such that  $\mu''(Y) = 2$  and  $\mu(Z) = 3$  will not do, because  $2 \not\geq 3$ . The crucial feature of ordinal scales is that nothing matters beyond order; which is why, if one defines scales by its admissible measure functions, ordinal scales are very liberal.

Might the scale of *good* be merely ordinal? Lassiter's answer (2017, p.177), with which I agree, is negative: the scale of *good* cannot be merely ordinal, because in addition to order, the distance between elements also matters.

The crucial data point here is the admissibility of quantificational adverbs such as *much*, *a little* or *a lot*. Note that there is an interpretative difference between the following two sentences:

- (7) Volunteering is better than donating.
- (8) Volunteering is much better than donating.

However vague and variable the meaning of *much* may be, the fact that one can imagine a situation in which (7) would be true while (8) false suggests that they do not mean the same thing.

- (9) Volunteering is better than donating, but not much better.

Informally, the admissibility of such modifiers imposes the requirement on the scale of *good* that the distance between measures be meaningful: (9) says that the value of volunteering is higher than the value of donating, but that the distance between them is not "much". If the scale of *better* were merely ordinal, all measure functions that respect the ordering between the two *relata* should be acceptable. *A fortiori*, measure functions according to which the difference in value between volunteering and donating amounts to "much" and measure functions according to which it does not should all be acceptable. But if that were so, it would not be possible to represent the contrast in truth conditions between (7) and (8). (9) attests such a contrast, and therefore the scale of better cannot be just ordinal.

More formally, the reasoning is the following: if *good* had an ordinal scale, then for any two elements on that scale that are ordered with respect to each other, all measure functions that respect that ordering should be

admissible. A sentence like (9) however, admits certain order-preserving measure functions but also rules out others, namely those that assign a value to each element that is at least as great as whatever quantity *much* stands for:

$$(10) \quad (9) = \mu(\textit{volunteer}) > \mu(\textit{donate}) \ \& \ [\mu(\textit{volunteer}) - \mu(\textit{donate})] \geq / \textit{much}$$

The fact that the truth-conditions of (9) require ruling out certain order-preserving measure functions suggests that the scale of *good* must have more structure than that of an ordinal scale. Regardless of how one defines *much*, there will be order-preserving measure functions for which the relation in (7) holds, but the one in (8) doesn't – just think of any measure function assigning some but not much difference in value to volunteering and donating.

Based on this, one can conclude that *good* must have a stronger scale than ordinal. The reader can check that similar observations apply to other evaluative adjectives, as they can all be modified by quantificational adverbs such as *much*, *a little* or *a lot*. The other two salient alternatives are interval and ratio scales, in order of increasing strength.

#### 4. Interval scales

Interval scales are stronger than ordinal scales, but weaker than ratio scales. They are stronger than ordinal scales because over and above mere order, the distance between elements on the scale, that is, their intervals, matters. However, they are weaker than ratio scales, because they do not determine a zero point, and therefore the positions of elements on the scale cannot be defined using ratios. Interval scales take into account the distance, or gaps, between elements – for this reason, the elements on an interval scale are not actually points, but intervals (although this will not matter for our purposes).

Temperature, clock time or danger are familiar examples of interval scales. Informally, what is crucial about those natural language cases is that the scales that those expressions lexicalize do not determine a zero point: a “zero” degree of temperature is a mere convention, and changes when one

moves from the Celsius to the Fahrenheit scale. Similarly, it is intuitive to think that there is no zero point in clock time or in a scale of danger/safety. Formally, this is cashed out by making ratio transformations meaningful only relative to some arbitrary reference point:

**Definition 2** (Interval Scale). *Where  $\mathcal{S} = \langle X, \succcurlyeq, \circ \rangle$  is a scale, if  $\mathcal{S}$  is an INTERVAL SCALE, then the following representation theorem holds for every admissible measure function  $\mu$  that maps  $\mathcal{S}$  onto  $\langle \mathbb{R}, \geq, + \rangle$ : for all*

*$x, y \in X$  and  $x \succcurlyeq y$ ,*

(i)  *$\mu(x) \geq \mu(y)$  and*

(ii) *for any  $\mu'$  satisfying condition (i) and for any  $z \in X$ , there are some  $n, m$  such that  $n \in \mathbb{R}^+$  and  $m \in \mathbb{R}$ ,  $\mu'(z) = n\mu(z) + m$ .*

That interval scales are strictly stronger than ordinal scales is easily seen by considering that the set of admissible measure functions according to **Definition 2** is a proper subset of the admissible measure functions according to **Definition 1**.

The crucial linguistic prediction associated with interval scales is that ratio modifiers are unacceptable. Recall the following examples:

- (5) a. # Bowl A is 1.38x as hot as bowl B.  
 b. # Amir came 2x as late as Mora.  
 c. # My neighborhood is 4x as safe as yours.

Those ratio comparisons are meaningless unless a zero point is defined on the relevant scale, but the adjective does not provide one.

So, an attractive explanation for why (5a) is odd becomes available: note that temperature is measured by scales such as Celsius, Fahrenheit or Kelvin. Now, (5a) might be true in a certain scale (say, Celsius). But if one moves to a Fahrenheit scale, the ratio 1.38 will be meaningless because the conversion between Celsius and Fahrenheit does not preserve ratios. For instance, if bowl A is 62.1°C and bowl B is 45°C, one could say that bowl A is 1.38x hotter on the Celsius scale than bowl B. But in a Fahrenheit scale, those temperatures are 143.78 and 113 respectively, and the ratio between them would no longer be 1.38.

However, I noted that (5a) can be repaired by mentioning a specific scale: if one adds the qualification that one is using a Celsius scale, the sentence immediately improves.

- (6) ✓ Bowl A is 1.38x as hot as bowl B on the Celsius scale.

The reason for this is that mentioning the Celsius scale introduces the necessary zero point required to interpret the ratio modifier.

Note, in addition, that the comparative size of intervals can be measured using ratios. So even though it does not make sense to say that Amir came twice as late as Mora, it does make sense to say that Amir was delayed by twice as much, or that he stayed for twice as long as Mora. This is because, even though the scale of temporal instants does not have a natural zero point, the intervals between temporal instants do.

Might *good* have an interval scale? The answer is not straightforward. On the one hand, *good* (and evaluatives in general) eschews precise ratio modifiers such as 1.38x.<sup>7</sup> In this sense *good* behaves like interval adjectives:

- (11) # Volunteering is 1.38x as good as donating.

However, this is not enough to conclude that *good* has an interval scale, for two reasons. First, attesting the unacceptability of ratio modifiers is not enough to determine that the relevant adjective has an interval scale. Postulating an interval scale is appropriate in the case of temperature or clock time, but that is because we know independently how temperature and clock time are measured – and, in particular, we know that zero points on the relevant scales are arbitrary. This reasoning does not apply to *good*: we do not know whether putative zero points on the goodness scale are arbitrary, because – again, setting aside coercive interpretations – there is no standard way of measuring value. Given that we lack independent evidence for or against the presence of arbitrary zero points on the *good* scale, we cannot conclude from the unacceptability of precise ratio modifiers that *good* has an interval scale.

Secondly, *good* (and other evaluatives) admit round ratio modifiers, whereas interval adjectives reject them:

- (12) You have to be twice as good.<sup>8</sup>

---

<sup>7</sup> Recall that uses in which *good* is coerced into a ratio interpretation are set aside, see n.3.

<sup>8</sup> Adapted from Coates' quote at the start.

- (13) Your daughter is, like, four times more beautiful.<sup>9</sup>  
 (14) He'd have to be ten times more charming than Arnold.<sup>10</sup>

Given this, the same reasoning regarding quantificational adverbs applies here. If some ratio modifiers are acceptable, this means that there is an interpretative difference between, e.g., *better* and *twice as good*.

- (15) You have to be better than Concha.  
 (16) You have to be twice as good as Concha.

If so, then the scale of *good* must be capable of representing this difference. However indeterminate the meaning of *twice as good* may be, the fact that one can imagine a situation in which (15) would be true while (16) is false suggests that they do not mean the same thing.

- (17) You have to be better, but not twice as good as Concha.

As discussed above, if an adjective accepts a modifier, then one can conclude that the adjective has a scale at least as strong as to represent the information contributed by the modifier. The admissibility of ratio modifiers imposes the requirement on the scale of *good* that ratios be meaningful. If the scale of *good* were merely interval, then it would not be possible to represent ratios between degrees. But (17) does represent a ratio between value measures, and therefore the scale of *good* cannot be simply interval. However, Lassiter (2017, 89 and ff) has resisted the view that the acceptability of round ratio modifiers is evidence against adjectives like *good* having an interval scale. In his view, round ratio modifiers are hyperbolic and stand for interval modifiers such as *much* or *a lot*. E.g., *ten times more charming* would be a hyperbolic way of saying *much more charming*. Lassiter says that the fact that those sentences become unacceptable when one adds an adverb like *exactly* points in this direction:

- (18) You have to be (# exactly) twice as good as Concha.  
 (19) Your daughter is, like, (# exactly) four times more beautiful.  
 (20) He'd have to be (# exactly) ten times more charming than Arnold.

<sup>9</sup> Adapted from the series *Fresh Off the Boat*, season 5 chapter 5, 2018.

<sup>10</sup> Adapted from the movie *Pulp Fiction*, 1994.



But it is one thing to say that these modifiers are hyperbolic, and a different one to say that they are really interval. The latter view predicts that such modifiers should be admissible with interval adjectives across the board. This prediction is not borne out, as *twice as hot* is just as bad as *1.38x hotter* (the same goes for (5b)-(5c), barring possible acceptable instances, see n.6):

(21) # Bowl A is twice as hot as bowl B.

To avoid this bad prediction, one may reject Lassiter's exact version of a hyperbole view, according to which round ratio modifiers are tantamount to interval modifiers. Alternatively, one could say that modifiers like *10x*, *20x* or *50x* are hyperbolic ways of saying *many times*. This possibility is suggestive when one considers the relative frequency of these modifiers in corpora: briefly, "very" round modifiers such as *2x*, *10x*, *100x* and *1000x* are significantly more frequent than *3x*, *4x*, *5x*, *20x* or *50x*. This suggests that the former might be somewhat idiomatic, and not to be taken as literally expressing measurement.<sup>11</sup> But even so, if *good* accepts a modifier like *many times*, this is still evidence that *good* has a stronger scale than interval, contrary to Lassiter.

In sum: partially based on his view that round ratio modifiers are hyperbolic, Lassiter maintains that *good* has an interval scale. But his reasoning is essentially abductive: given that according to him an ordinal and a ratio scale can be ruled out, only interval scales remain as a candidate among the type of scales attested in natural language. I have offered an argument against the view that *good* has an ordinal scale (acceptability of *much*); as well as an argument against the view that *good* has an interval scale (acceptability of *twice*). Lassiter rejects the latter, but I've pushed back against his alternative view that round ratio modifiers are hyperbolic

---

<sup>11</sup> A search on Corpus of Contemporary American English (COCA, <https://www.english-corpora.org/coca/>) of a set of round ratio modifiers between 1 and 100, in addition to 1000, combined with *better* reveals that a handful of round modifiers are significantly more frequent than others. In decreasing frequency: *twice as good* (113), *ten times* (84), *a thousand times* (63), *a hundred times* (38), *five times* (18), *three/four times* (15), *twenty times* (4) and *fifty times better* (3). I thank a reviewer for inviting me to look into this.

interval modifiers. However, I haven't yet looked at Lassiter's argument against *good* having a ratio scale. To this end, let us move on to ratio scales.

## 5. Ratio scales

Ratio scales are characterized by the fact that the relative “size” of elements matters. In particular, difference in size between elements is measured in ratios, which means that only ordering-preserving measure functions that are obtained via a multiplication operation are admissible. In addition to this, ratio scales require that concatenation be mapped onto the mathematical operation of addition. That is, the concatenation of two elements may only be mapped onto a measure function that assigns to such compound object the arithmetical sum of the individual measures of the concatenated elements.

Scales like height and weight are familiar examples of ratio scales, where the relation between elements in the scale can be mapped onto measure functions that maintain a constant ratio between the numerical values assigned to them. More formally:

**Definition 3** (Ratio Scale). *If a scalar property  $\mathcal{S} = \langle X, \succcurlyeq, \circ \rangle$  is a RATIO SCALE, then the following representation theorem holds for every admissible measure function  $\mu$  that maps  $\mathcal{S}$  onto  $\langle \mathbb{R}, \geq, + \rangle$ : for all  $x, y \in X$  and  $x \succcurlyeq y$ ,*

(i)  $\mu(x) \geq \mu(y)$ ,

(ii)  $\mu(x \circ y) = \mu(x) + \mu(y)$  and

(iii) for any  $\mu'$  satisfying (i) and (ii), there's an  $n \in \mathbb{R}^+$  s.t. for any  $z \in X$ ,  $\mu'(z) = n\mu(z)$ .

Recall that admissible measure functions for ordinal scales satisfy only the first of those conditions. Admissible measure functions for interval scales satisfy the first condition as well as a “liberal” version of the third, where ratios are calculated relative to arbitrary and variable “zero” points. Since no reference is made here to such variable, ratios are fixed relative to the real zero. Thus, whereas an interval scale admits all ratios calculated taking any real as reference point, a ratio scale admits only those calculated relative to 0. This means that a ratio scale imposes more conditions on the

admissible measure functions, and is therefore a stronger scale type than ordinal and interval scales.

In order to see how ratio scales constrain admissible measure functions, consider a familiar example: height. Seeing why the height scale  $\mathcal{S}_{height}$  is stronger than an ordinal scale is straightforward: suppose that Amir is taller than Mora. If height were an ordinal scale, one should be able to map Amir and Mora's heights to any pair of numerical values under the  $>$  relation. But some of those values would radically misrepresent their heights. Suppose that Amir and Mora are respectively 182 and 165 centimeters tall. Consider a measure function  $\mu'$  that assigns  $\mu'(Amir) = 182$ ,  $\mu'(Mora) = 165$ , but such that their concatenated heights,  $\mu'(Amir \circ Mora)$ , is equal to 17.  $\mu'$  respects the ordering relation between them – i.e., complies with condition (i) in **Definition 3**. But it radically misrepresents the intuitive value of their concatenated heights – it doesn't comply with condition (ii). Or consider another measure function  $\mu''$  that assigns  $\mu''(Amir) = 182$ ,  $\mu''(Mora) = 181.9$  and  $\mu''(Amir \circ Mora) = 363.9$ . This measure function respects the ordering relation between Amir and Mora – complying with (i), and the fact that their combined heights should be the arithmetical sum of their individual heights – complying with (ii). But it does not respect the intuitive relation that holds between Amir and Mora's heights, because it does not preserve the ratio between their heights. That is, it does not respect condition (iii) in **Definition 3**.

Condition (ii) and (iii) in **Definition 3** impose more structure on the admissible measure functions for a ratio scale than mere preservation of order, and thereby define a stronger scale. In particular, if a scale  $\mathcal{S}$  is ratio, only order-, addition-, and ratio-preserving measure functions are admissible.

I have argued that *good* cannot have an ordinal scale, and I have argued that it doesn't have an interval scale either. Is the goodness scale a ratio scale? There are two considerations against this. First, since ratio scales make ratio comparisons interpretable, adjectives that have a ratio scale are predicted to admit ratio modifiers. This prediction is borne out for *tall*, which is (independently) known to have a ratio scale:

$$(22) \quad \text{Amir is 1.38x as tall as Mora.}$$

Conversely, adjectives that eschew ratio modifiers are predicted to not have ratio scales. Such is the case for *good* and precise ratio modifiers:

- (11) # Volunteering is 1.38x as good as donating.

Secondly, ratio scales are by definition POSITIVE with respect to concatenation, and the scale of *good* is not, according to Lassiter (2017, p. 179 and ff). Being positive with respect to concatenation means that the concatenation of any two elements has a greater degree of the relevant property than either element. More formally, a scale  $\mathcal{S} = \langle X, \geq, \circ \rangle$  is positive with respect to concatenation iff for any  $x, y \in X$  that do not overlap,  $x \circ y > x$  (except if  $\mathcal{S}$  is lower-bounded, and  $y$  has exactly the value of the lower-bound; i.e., if  $y$  is equal to 0). Lassiter holds that the *good* scale lacks this property, based on the observation that it seems to respect the following inference pattern:

- (23) a.  $a \geq b$   
 b.  $a \geq c$   
 $\therefore a \geq (b \circ c)$

If  $\mathcal{S}$  were positive with respect to concatenation, that inference should fail in many instances. But it does not fail for *good* (by contrast, it very clearly fails for *likely*, which is independently argued to have a ratio scale). For an example, consider the following, intuitively valid inference from Lassiter (2017, p. 179; recall that concatenation for propositions is disjunction):

- (24) a. It's as good for the card to be a spade as it is for it to be a heart.  
 b. It's as good for the card to be a spade as it is for it to be a diamond.  
 $\therefore$  It's as good for the card to be a spade as it is for it to be a red card.

According to Lassiter, that this inference pattern is in general valid shows that the scale of *good* has to be weaker than a ratio scale.

Here appears a dilemma. The first horn is that considerations about precise ratio modifiers suggest rejecting all scale types as weak as ratio. The second horn is that considerations about concatenation suggest rejecting any scale type as strong as ratio, since all are positive with respect to

concatenation. I propose to solve this dilemma by partially rejecting the second horn: Lassiter's considerations about concatenation apply to propositional-level *good* (*it is good that p*), but not to individual-level, or adnominal *good* (*x is good*). Thus, even though there's reason to reject any scale type as strong as ratio for propositional *good*, those considerations do not extend to individual-level *good*. In sum, I agree with Lassiter that propositional-level *good* has an interval scale, but I'll propose that individual-level *good* has a scale that is stronger than interval, although weaker than a standard ratio scale.

When one moves from propositional to adnominal *good*, the inference in (24) arguably fails. Intuitively, this is the case because concatenation for individuals is mereological sum, and the sum of two individuals can have a higher value than each of those individuals taken separately:<sup>12</sup>

- (25) a. Car *a* is at least as good as car *b*.  
 b. Car *a* is at least as good as car *c*.  
 $\therefore$  Car *a* is at least as good as car  $b \oplus c$ .

To see how this inference can fail, one can think of good in terms of preference: for any two individuals *x, y*, *x* is at least as good as *y* just in case *x* is at least as preferable as *y*. In turn, one may spell this out by saying that *x* is at least as preferable as *y* just in case every time you have the option of choosing *y*, you also choose *x*. Understood in this way, premise (25a) says that every time you have the option of choosing *b*, you also choose *a*. Premise (25b) says that every time you have the option of choosing *c*, you also choose *a*. But it is consistent with this that if you get to choose the sum of *b* and *c*, you may no longer choose *a* as well. In other words, cars *b* and *c* may have a higher value taken together than taken separately, making the premises true but the conclusion false.

The key to the contrast between adnominal and propositional-level *good* is, of course, concatenation: concatenation for individuals is mereological sum, while for propositions it is disjunction. This has completely different

---

<sup>12</sup> One might disagree about specific cases, perhaps with other evaluative adjectives. Can the sum of two dishes be tastier than the tastiest of them? Can the sum of two pictures be more beautiful than the most beautiful of the two? Perhaps not, but a single positive instance suffices to falsify the inference, and it can be found.

consequences for the assessment of complex objects. Note that, intuitively, the value of a proposition amounts to the value of its outcome. Similarly, the value of a disjunction must also amount to the value of its outcome, that is, one of its disjuncts. This suggests that the value of a disjunction is maximal, that is, a disjunction is no more valuable than its most valuable disjunct. By contrast, the value of a mereological sum of individuals is potentially positive, that is, higher than the value of each individual in it.

Lassiter relies on examples like (24) to conclude that the scale of propositional *good* can't be ratio, and I agree with his conclusion. But (25) shows that this does not extend to individual *good*. This suggests that propositional and individual good might have different scales. In the next section I will argue that a further observation supports this conclusion, and I will propose a stronger scale type for individual-level *good* than for propositional-level *good*.

## 6. Round ratio scales

In this last section, I want to propose that individual-level *good* has a stronger scale than interval, but not as strong as a standard ratio scale (as defined in **Definition 3**). Informally, the idea is the following: whereas a standard ratio scale requires that admissible measure functions preserve a precise ratio, which is a positive real, one can define a type of ratio scale according to which this requirement is relaxed, so that admissible measure functions preserve only an approximate ratio. I will call this type of scale a ROUND RATIO SCALE. In practice, this means that a round ratio scale rules out less measure functions than a standard ratio scale, and is thereby weaker.

To define such a scale, one can impose the requirement that the ratio that gets preserved across measures of the same individual is not a positive real, but some positive in its vicinity, defined by a HALO. A halo is an interval around a number whose size can vary. For example, the halo of 2 could be the interval  $[1.9, 2.1]$ , or  $[1.8, 2.2]$ .<sup>13</sup>

---

<sup>13</sup> See Lasersohn 1999 on how halos are at play in the interpretation of numerical expressions, as well as Hoek 2018; Sauerland and Stateva 2011 for elaboration and criticism of Lasersohn's seminal view.

It is well-known that rounder numbers have greater halos; e.g., 10 has a greater halo than 11 or 9; 50 has a greater halo than 49 or 51.<sup>14</sup> But what numbers are round, and why do they have greater halos? Defining round numbers is not as straightforward as it may seem; for my purposes, I will rely on the following informal and comparative definition: an integer is round just in case it has a larger number of smaller factors than its neighboring numbers (Hardy 1940, p. 48). For example, 10 has more and smaller factors than 9 and 11; the same goes for 50 as opposed to 49 and 51. Even more informally, one tends to consider rounder numbers that end on one or more zeros (relative to a given base), as well as simple multiples or fractions of such numbers (Sigurd 1988, p. 249).

Regarding the question of why rounder numbers have greater halos, one possible answer is that round numbers are cognitively significant (see Rosch 1975, who characterizes round numbers as a kind of *cognitive reference points*). Alternatively, or perhaps as a result of their cognitive significance, round numbers tend to be linguistically simpler (Lotz 1955, see also Krifka 2002, 2007). The cognitive significance of round numbers, sometimes called the *round number bias*, has been studied in domains such as psychology or economics (Lacetera et al. 2012; Lynn et al. 2013; Pope and Simonsohn 2011).<sup>15</sup>

Having characterized halos and round numbers, let's now define a round ratio scale. The definition is similar to the standard ratio scale, except that condition (iii) in **Definition 3** above is relaxed, so that admissible ratio transformations are restricted, not to those that preserve some real, but to those that preserve some real within some other real's halo. This is achieved

---

<sup>14</sup> This is reflected in loose talk. To take an example from Hoek (2018, p. 175), at 3:58 it is preferable to say *4 o' clock* than *3:57*, even though the latter is closer to the truth.

<sup>15</sup> A salient manifestation of the round number bias is the *left-digit effect*, which explains the tendency to price items right below round numbers, such as 3.99€. Buyers perceive the difference between 3.99 and 4 as more meaningful than the difference between, e.g., 4 and 4.01, and sellers take advantage of it (see Bhattacharya et al. 2012). Another interesting manifestation of round number bias is the strive, in sports and other domains, to attain round scores (Lotz 1955; Pope and Simonsohn 2011).

by substituting a fixed ratio for a *halo function*. A halo function  $H_k$  is a function from  $\mathbb{R}^+$  to  $\mathbb{R}^+$  such that, for any  $m, n, k \in \mathbb{R}^+$ ,  $H_k(n) = m$  just in case  $m$  is the result of some simple arithmetical operation on  $n$  that maps  $n$  onto some number that is no further from  $n$  than the halo size of  $k$ . For example, suppose that  $k = 2$ . Assuming that the halo of 2 is the interval  $[1.9 - 2.1]$ , whose size is 0.2, there's infinitely many functions  $H_2$ , all those functions that take their argument  $n$  to any number no further away from  $n$  than 0.2. Here are some examples of possible functions  $H_2$  (note that this includes the identity function):

- $H_2^i = \lambda n. n$
- $H_2^j = \lambda n. n + 0.1$
- $H_2^k = \lambda n. n - 0.05$
- ...

But note that, e.g., a function  $\lambda n. n + 0.3$  is not such a function, as it maps its argument further away from the halo size of 2.

If, instead of imposing the requirement that admissible transformations preserve some ratio  $n$ , one imposes the requirement that they preserve a ratio that results from mapping  $n$  to some number in its vicinity, one can allow the necessary variability. Let us define Round Ratio Scales as follows:

**Definition 4** (Round Ratio Scale). *If a scalar property  $\mathcal{S} = \langle X, \geq, \circ \rangle$  is a ROUND RATIO SCALE, then the following representation theorem holds for every admissible measure function  $\mu$  that maps  $\mathcal{S}$  onto  $\langle \mathbb{R}, \geq, + \rangle$ : for all  $x, y \in X$  and  $x \geq y$ ,*

(i)  $\mu(x) \geq \mu(y)$ ,

(ii)  $\mu(x \circ y) = \mu(x) + \mu(y)$  and

(iii) for any  $\mu'$  satisfying (i) and (ii), there are  $n, m \in \mathbb{R}^+$  s.t., for any  $z \in X$ , there's some function  $H_m$  s.t.  $\mu'(z) = H_m(n)\mu(z)$ .

A round ratio scale does not require that the ratios between the measures assigned to each individual are held constant across admissible measure functions; rather, such ratios are allowed to vary within a certain halo. Thus, for instance, given an admissible measure function  $\mu$  such that  $\mu(x) = 2$  and  $\mu(y) = 1$  (for any  $x, y \in X$  such that  $x \geq y$ ), consider a measure func-



tion  $\mu'$  such that  $\mu'(x) = 4.1$  and  $\mu'(y) = 2$ . Given  $\mu$ ,  $\mu'$  would be inadmissible in a standard ratio scale, since there is no positive real  $n$  such that, for every  $z \in X$ ,  $\mu'(z) = n\mu(z)$ . For  $x$ ,  $\mu'(x)/\mu(x)$  is 2.05; while for  $y$ ,  $\mu'(y)/\mu(y)$  is 2. Those ratios are not the same, and *a fortiori* there does not exist a single ratio for all measure functions applied across all elements of  $X$ .

But  $\mu'$  is an admissible measure function according to **Definition 4**. The reason is that, even though  $\mu'(x)/\mu(x) \neq \mu'(y)/\mu(y)$ , that is,  $2.05 \neq 2$ , there exist halo functions  $H_m$ , for some positive real  $m$ , such that one can map either of these ratios,  $\mu'(x)/\mu(x)$  or  $\mu'(y)/\mu(y)$ , to match some positive real, namely 2. First, consider  $x$ . There is a function  $H_m$  such that  $\mu'(x) = H_m(2)\mu(x)$ . This function, call it  $H_m^i$ , is  $\lambda n.n + 0.05$ . Substituting  $H_m$  for  $H_m^i$  in  $\mu'(x) = H_m(2)\mu(x)$ , we obtain  $4.1 = \lambda n.n + 0.05(2) \times 2$ , that is,  $4.1 = 2 + 0.05 \times 2$ , which is true. Secondly, consider  $y$ . There is also a function such that  $\mu'(y) = H_m(2)\mu(y)$ . This function, call it  $H_m^j$ , is just the identity function,  $\lambda n.n$ . Substituting  $H_m$  for  $H_m^j$  in  $\mu'(y) = H_m(2)\mu(y)$ , we obtain  $2 = \lambda n.n(2) \times 1$ , that is,  $2 = 2 \times 1$ , which is true.

This opens up the possibility that ratios are calculated only approximately. However, here appears a hurdle: given that the set of reals  $\mathbb{R}^+$  is countably infinite, halos can be of countably infinite size as well. This means that **Definition 4** does not, after all, rule out any measure function (beyond those that fail conditions (i) or (ii)): however different the ratio assigned by two measure functions to a pair of individuals may be, their difference will fall within the halo of *some* real. For example, suppose again that an admissible measure function  $\mu$  is such that  $\mu(x) = 2$  and  $\mu(y) = 1$ . According to **Definition 4**, any other admissible measure function  $\mu'$  must be such that there exist  $n, m \in \mathbb{R}^+$  such that, for every  $z \in X$ , there exists some function  $H_m$  such that  $\mu'(z) = H_m(n)\mu(z)$ . The issue is that there will always be some positive real  $m$  whose halo is as great as required, so there is in fact no restriction on how far ratios can come apart. In sum, **Definition 4** above is too weak.

Regardless, one can adopt **Definition 4** as a template, and use it to define different round ratio scales of specific granularity. By assigning a specific granularity, one determines a maximum halo size that ratio transformations are allowed to vary within, thereby restricting the admissible measure functions in a way that strengthens Definition 4:

**Definition 5** (Round Ratio Scale of  $n$ -granularity). *If a scalar property  $\mathcal{S} = \langle X, \succcurlyeq, \circ \rangle$  is a ROUND RATIO SCALE OF  $n$ -GRANULARITY, then the following representation theorem holds for every admissible measure function  $\mu$  that maps  $\mathcal{S}$  onto  $\langle \mathbb{R}, \geq, + \rangle$ : for all  $x, y \in X$  and  $x \succcurlyeq y$ ,*

(i)  $\mu(x) \geq \mu(y)$ ,

(ii)  $\mu(x \circ y) = \mu(x) + \mu(y)$  and

(iii) for any  $\mu'$  satisfying (i) and (ii), there's an  $m \in \mathbb{R}^+$  s.t., for any  $z \in X$ , there's some function  $H_n$  s.t.  $\mu'(z) = H_n(m)\mu(z)$ .

Thus, for example, if a scalar property has a round ratio scale of 2-granularity, then the absolute difference between the ratios assigned by any two admissible measure functions to any pair of individuals cannot be greater than  $H(2)$ , that is,  $[1.9 - 2.1] = 0.2$ . Recall  $\mu$  and  $\mu'$ . Given the measures assigned to  $x, y \in X$ ,  $\mu'$  was an admissible measure function according to **Definition 4**. But what about according to **Definition 5**? It depends on whether, for every  $z \in X$ , there's a function  $H_2$  – a function that maps its argument to a number no further away from it than 0.2 – such that  $\mu'(z) = H_2(m)\mu(z)$ , for some positive real  $m$ . Such functions were already found for  $x$  and  $y$ , namely  $H_m^i = \lambda n.n + 0.05$  and  $H_m^j = \lambda n.n$ , respectively. Therefore, given  $\mu, \mu'$  is an admissible function for a round ratio scale of granularity 2.

But consider, by contrast, another measure function  $\mu''$  according to which  $\mu''(x) = 6$  and  $\mu''(y) = 2$ . Is there some positive real  $m$  such that, for every  $z \in X$ , there is some function  $H_2$  such that  $\mu''(z) = H_2(m)\mu(z)$ ? Recall that it's not necessary for  $\mu''(x)/\mu(x)$  to be identical to  $\mu''(y)/\mu(y)$ , as would be the case in a standard ratio scale. Rather, what is needed is for those ratios to vary at most by 0.2. In other words, the absolute difference between  $\mu''(x)/\mu(x)$  and  $\mu''(y)/\mu(y)$  has to not be greater than 0.2. But since  $\mu''(x)/\mu(x) = 3$  and  $\mu''(y)/\mu(y) = 2$ , this is not the case. Thus, given  $\mu, \mu''$  is not an admissible measure function of a round ratio scale of granularity 2.

Importantly, recall that rounder numbers have greater halos. Therefore, the “rounder” its granularity, the weaker a round ratio scale will be: a round ratio scale of 20-granularity will be weaker than one of 2-granularity, which will be weaker than one of 0.2-granularity, and so on. This means that a measure function that is not admissible on a given round ratio scale might

be admissible on a round ratio scale with rounder granularity. Suppose that the halo of 5 is  $[4.5 - 5.5]$ , that is, 1. As just shown, given  $\mu$ ,  $\mu''$  above would not be an admissible measure function of a round ratio scale of granularity 2. But  $\mu''$  would be an admissible measure function of a round ratio scale of granularity 5. For that, the absolute difference between  $\mu''(x)/\mu(x)$  and  $\mu''(y)/\mu(y)$  need not be greater than the halo size of 5, that is, 1. And it isn't. Therefore, bigger halos make for weaker scales.

The proposal is, then, that adnominal *good* has a round ratio scale of  $n$ -granularity. But what  $n$ ? Settling this requires saying something about ratio modifiers. A simple way of cashing out the meaning of any ratio modifier is to assign to it the presupposition that the adjective with which it combines has a ratio scale, and then assign to it the at-issue meaning one would expect:

- (26)  $[[x \text{ is } n\text{-}x \text{ as } A \text{ as } y]] = \text{defined only if } A \text{ has a ratio scale.}$   
 If so,  $[[x \text{ is } n\text{-}x \text{ as } A \text{ as } y]] = 1 \text{ iff } \mu_A(x) = n\mu_A(y)$

According to this simple proposal, ratio modifiers should be acceptable across the board with ratio adjectives. This prediction is borne out for standard ratio adjectives such as *tall*, as they admit any ratio modifier. But it fails for evaluative adjectives such as *good*, which admit some ratio modifiers (2x) but not others (1.38x).

However, if one modifies the presupposition of ratio modifiers, so that their number indicates the granularity of the ratio scale that they require, their acceptability can serve as a guide to the granularity of ratio adjectives. Here is a proposal:

- (27)  $[[x \text{ is } n\text{-}x \text{ as } A \text{ as } y]] = \text{defined only if } A \text{ has a ratio scale of } n\text{-granularity.}$  If so,  $[[x \text{ is } n\text{-}x \text{ as } A \text{ as } y]] = 1 \text{ iff } \mu_A(x) = n\mu_A(y)$

According to this view, ratio adjectives have the granularity of the most precise ratio modifier that they accept. Standard ratio adjectives, like *tall*, have maximal granularity, and thus accept any ratio modifier. But other adjectives have less-than-maximal granularity. Adnominal *good*, for example, will be of 2-granularity, since this is likely the most precise ratio modifier that it can admit. More precise ratio modifiers, such as *1.38x*, will require a ratio scale of *1.38-granularity*, which is too precise for *good*, and this is why a phrase like *1.38x better* is infelicitous. In turn, since according

to (27) ratio modifiers presuppose that their adjective has at least their granularity, *m-x as A* will be infelicitous if *A* does not have a ratio scale of at least *m*-granularity.

Moreover, I argued that adnominal *good* is positive with respect to concatenation (cf. (25)). Round ratio scales are so as well, so this prediction is also borne out. In sum, I am claiming that *good*, when taking individuals as its arguments, has a round ratio scale, which is a type of scale that is weaker than standard ratio insofar as it admits more measure functions than standard ratio scales.

Before moving on, an important question remains: why would evaluative adjectives in general, and adnominal *good* in particular, have scales that blur precise ratios? Relying on broad views about vagueness, one may distinguish three genres of response: metaphysical – because there’s no such thing as precise ratios of value, epistemic – because we cannot know objectively where precise ratios are, and psychological – because we are psychologically insensitive to them. Setting aside a metaphysical view, which would require much more discussion than I have space for, an intuitive justification for going in for an epistemic view would be, perhaps, that we simply haven’t yet figured out how to measure value precisely enough in an intersubjectively verifiable way. That is, it may be only subjectively possible to distinguish *2x as good* from *1.9x* or *2.1x* as good. Intersubjective measures of value might just be approximate. This view might be bolstered by the fact, pointed out in the Introduction, that *good* is a judge-dependent and/or multidimensional predicate. The idea could be that, even though each of us may be able to subjectively determine a standard ratio scale of value – perhaps through some operation of dimension aggregation – the best we can do to share such measurements with others are rough approximations, that nevertheless succeed in preserving the overall scalar architecture.<sup>16</sup>

A psychological view, on the other hand, might be supported by features of our perceptual and cognitive system. It is well-known that, even though we are capable of representing magnitudes in a mathematically precise way, our perceptual system represents magnitudes in an analog fashion, assigning

---

<sup>16</sup> I thank two reviewers for this journal for independently pointing to this hypothesis. See also Sassoon 2010, p. 161 and ff., for ideas in this vicinity.

measures to objects that, even though globally covariant with the represented magnitude, introduce a great deal of probabilistic error.<sup>17</sup> It is not wholly counterintuitive to think that, even though the linguistic expression of many such magnitudes (height, weight, distance) inherit their scalar properties from our mathematical capacity to represent precise magnitudes, expressions that denote the *value* of such magnitudes, that is, evaluative adjectives, inherit their scalar properties from our imprecise perceptual system.

Discussing these hypotheses further will have to wait for another occasion. For now, I take it to be at least a competing hypothesis that adnominal *good* has a round ratio scale. By contrast, I conclude with Lassiter that propositional-level *good* has an interval scale. To further support this, note the following: whereas a phrase such as *twice as good* is fairly common, it is hardly ever used to compare the value of propositions. Examples do not abound,<sup>18</sup> and when they are felicitous, they seem to inherit their acceptability from an individual-level comparison. For example, one can say something like:

- (28) It is good that Camila came to the party. It would have been twice as good if she had come with Milica!

Even though this sentence compares the value of propositions, the aggregated value of Camila and Milica coming to the party cannot result from the concatenation of the proposition that Camila comes to the party and the proposition that Milica comes to the party, since the concatenation of those two propositions is their disjunction. Somehow, the aggregated value of Camila and Milica coming to the party is vicariously calculated by aggregating their value as individuals.

Assuming that these examples are rare, and derive their meaning from an individual-level evaluation, the observation that ratio modifiers are unacceptable with propositional-level *good* is predicted if propositional-level *good* has an interval scale, since *twice* is too strong for propositional-level *good*.

---

<sup>17</sup> This observation is familiar from the literature on vagueness, see e.g., Égré 2017; Fults 2011.

<sup>18</sup> None of the 113 hits of *twice as good* at COCA apply to propositions.

## 7. Conclusion

In this paper, I have discussed the scale of evaluative adjectives, and, in particular, of the evaluative adjective *good*. I've argued, first, that the lack of measure phrases gives linguistic support to the (otherwise natural) view that *good* lacks measurement units. However, *good* does not eschew measurement altogether. I have argued that there are strong reasons to think that the scale of *good* is stronger than a mere ordinal scale; and I have moved on to discuss which of the two most salient candidates discussed in the literature is most appropriate for *good*, an interval or a ratio scale.

Interval scales were rejected based chiefly on the observation that evaluatives admit round ratio modifiers, but this conclusion was not free of controversy, as Lassiter has argued that those are hyperbolic uses. Ratio scales, by contrast, were partially rejected based on features of concatenation. Ratio scales are by definition positive with respect to concatenation, and it was observed that, while there is no evidence of propositional-level *good* being positive with respect to concatenation, one can argue that individual-level *good* is. Based on this, I proposed to sever the scales of propositional- and individual-level *good*: while the former has an interval scale, the latter has a ratio scale. More specifically, I've proposed that individual-level *good* has a ROUND RATIO SCALE, a type of scale that preserves approximate rather than precise ratios, and is thereby stronger than interval, but weaker than a standard ratio scale.

One potentially controversial aspect of this view remains to be discussed. What are the consequences of severing the scale of propositional- and individual-level *good*? As mentioned in the Introduction, this proposal is bound to be met with resistance, since there's *prima facie* reasons to maintain a uniform view about the scalar semantics of *good*. However, in light of the massive underspecificity of *good*, the prospects of this schism may seem less controversial. In fact, the view that "there's more than one *good*", that is, that *good* is ambiguous or polysemous, is perhaps not such a revisionary hypothesis in light of the properties of *good* reviewed in the Introduction. Indeed, the arguments put forward in this paper might be seen as supporting that general hypothesis. Moreover, the relationship between individual- and propositional-level *good* is understudied, so the view that these

evaluatives might have different scales is not at odds with any existing proposal that I know of. This divergence may simply be one more among other puzzling properties of evaluative adjectives.

### Acknowledgments

Special thanks to Erich Rast for multiple comments on various versions of this work. Thanks to Federico Faroldi, Mora Maldonado, Isidora Stojanovic, as well as to two anonymous reviewers for *Organon F* and Dan Zeman for his careful editorial work. This work has been funded by Portuguese national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., within project UIDB/00183/2020 and by COST Action CA17132, funded by the Horizon 2020 Framework Programme of the European Union.

### References

- Asher, Nicholas. 2011. *Lexical Meaning in Context: A Web of Words*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511793936>
- Bhattacharya, Utpal, Craig W. Holden, and Stacey Jacobsen. 2012. “Penny Wise, Dollar Foolish: Buy–sell Imbalances on and Around Round Numbers.” *Management Science* 58 (2): 413–31. <https://doi.org/10.1287/mnsc.1110.1364>
- Bylinina, Lisa. 2017. “Judge-dependence in Degree Constructions.” *Journal of Semantics* 34 (2): 291–331. <https://doi.org/10.1093/jos/ffw011>
- Castroviejo, Elena, and Berit Gehrke. 2019. “Intensification and Secondary Content: A Case Study of Catalan *Good*.” In *Secondary Content: The Semantics and Pragmatics of Side Issues*, edited by Daniel Gutzmann, and Katharina Turgay, 108–43. Brill. [https://doi.org/10.1163/9789004393127\\_006](https://doi.org/10.1163/9789004393127_006)
- Égré, Paul. 2017. “Vague Judgment: A Probabilistic Account.” *Synthese* 194 (10): 3837–65. <https://doi.org/10.1007/s11229-016-1092-2>
- Fults, Scott. 2011. “Vagueness and Scales.” In *Vagueness and Language Use*, edited by Paul Égré, and Nathan Klinedinst, 25–50. Springer.
- Geach, Peter T. 1956. “Good and Evil.” *Analysis* 17 (2): 33–42. <https://doi.org/10.2307/3326442>
- Hardy, Godfrey Harold. 1940. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work*. Cambridge University Press.
- Hare, Richard M. 1952. *The Language of Morals*. Oxford University Press. <https://doi.org/10.1093/0198810776.001.0001>
- Hoek, Daniel. 2018. “Conversational Exculpature.” *Philosophical Review* 127 (2): 151–96. <https://doi.org/10.1215/00318108-4326594>

- Krantz, David H, Patrick Suppes, and Robert Duncan Luce. 1971. *Foundations of Measurement*. New York Academic Press.
- Krifka, Manfred. 1989. "Nominal Reference, Temporal Constitution and Quantification in Event Semantics." In *Semantics and Contextual Expression*, edited by Renate Bartsch, Johan Van Benthem, and Peter Van Ende Boas, 75–115. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110877335-005>
- Krifka, Manfred. 2002. "Be Brief and Vague! And How Bidirectional Optimality Theory Allows for Verbosity and Precision." In *Sounds and Systems: Studies in Structure and Change. A Festschrift for Theo Vennemann*, edited by David Restle, and Dietmar Zaefferer, 439–58. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110894653.439>
- Krifka, Manfred. 2007. *Approximate Interpretation of Number Words*. Manuscript, Humboldt-Universität zu Berlin, Philosophische Fakultät II. <https://doi.org/10.18452/9508>
- Lacetera, Nicola, Devin G. Pope, and Justin R. Sydnor. 2012. "Heuristic Thinking and Limited Attention in the Car Market." *American Economic Review* 102 (5): 2206–36. <https://doi.org/10.1257/aer.102.5.2206>
- Lasersohn, Peter. 1999. "Pragmatic Lalos." *Language* 75 (3): 522–51. <https://doi.org/10.2307/417059>
- Lassiter, Daniel. 2017. *Graded Modality: Qualitative and Quantitative Perspectives*. Oxford: Oxford University Press.
- Lotz, John. 1955. "On Language and Culture." *International Journal of American Linguistics* 21 (2): 187–9. <https://doi.org/10.1086/464329>
- Lynn, Michael, Sean Masaki Flynn, and Chelsea Helion. 2013. "Do Consumers Prefer Round Prices? Evidence from Pay-What-You-Want Decisions and Self-Pumped Gasoline Purchases." *Journal of Economic Psychology* (36): 96–102. <https://doi.org/10.1016/j.joep.2013.01.010>
- Pope, Devin & Uri Simonsohn. 2011. "Round Numbers as Goals: Evidence From Baseball, SAT Takers, and the Lab." *Psychological Science* 22 (1): 71–9. <https://doi.org/10.1177%2F0956797610391098>
- Ridge, Michael. 2014. *Impassioned Belief*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199682669.001.0001>
- Rosch, Eleanor. 1975. "Cognitive Reference Points." *Cognitive Psychology* 7 (4): 532–47. [https://doi.org/10.1016/0010-0285\(75\)90021-3](https://doi.org/10.1016/0010-0285(75)90021-3)
- Sassoon, Galit W. 2013. "A Typology of Multidimensional Adjectives." *Journal of Semantics* 30 (3): 335–80. <https://doi.org/10.1093/jos/ffs012>
- Sassoon, Galit W. 2010. "Measurement Theory in Linguistics." *Synthese* 174 (1): 151–80. <https://doi.org/10.1007/s11229-009-9687-5>



- 
- Sauerland, Uli, and Penka Stateva. 2011. "Two Types of Vagueness." In *Vagueness and Language Use*, edited by Paul Egré, and Nathan Klinedinst, 121–45. Springer.
- Sigurd, Bengt. 1988. "Round Numbers." *Language in Society* 17 (2): 243–52. <https://doi.org/10.1017/S0047404500012781>
- Solt, Stephanie. 2018. "Multidimensionality, Subjectivity and Scales: Experimental Evidence." In *The Semantics of Gradability, Vagueness, and Scale Structure*, edited by Elena Castroviejo, Louise McNally, and Galit W. Sassoon, 59–91. Springer. [https://doi.org/10.1007/978-3-319-77791-7\\_3](https://doi.org/10.1007/978-3-319-77791-7_3)
- Stojanovic, Isidora. 2016. "Expressing Aesthetic Judgments in Context." *Inquiry* 59 (6): 1–23. <https://doi.org/10.1080/0020174X.2016.1208922>
- Thomson, Judith. 2008. *Normativity*. Open Court.
- Umbach, Carla. 2016. "Evaluative Propositions and Subjective Judgments." In *Subjective Meaning: Alternatives to Relativism*, edited by Cécile Meier, and Janneke van Wijnberger- Huitink, 127–68. Berlin: De Gruyter. <https://doi.org/10.1515/9783110402001-008>

# The Derogatory Force and the Offensiveness of Slurs

Chang Liu\*

Received: 13 November 2020 / Revised: May 3 2021 / Accepted: 5 June 2021

*Abstract:* Slurs are both derogatory and offensive, and they are said to exhibit “derogatory force” and “offensiveness.” Almost all theories of slurs, except the truth-conditional content theory and the invocational content theory, conflate these two features and use “derogatory force” and “offensiveness” interchangeably. This paper defends and explains the distinction between slurs’ derogatory force and offensiveness by fulfilling three goals. First, it distinguishes between slurs’ being derogatory and their being offensive with four arguments. For instance, ‘Monday’, a slur in the Bostonian argot, is used to secretly derogate African Americans without causing offense. Second, this paper points out that many theories of slurs run into problems because they conflate derogatory force with offensiveness. For example, the prohibition theory’s account of offensiveness in terms of prohibitions struggles to explain why ‘Monday’ is derogatory when it is not a prohibited word in English. Third, this paper offers a new explanation of this distinction from the perspective of a speech act theory of slurs; derogatory force is different from offensiveness because they arise from two different kinds of speech acts that slurs are used to perform, i.e., the illocutionary act of derogation and the perlocutionary act of offending. This new explanation avoids the problems faced by other theories.

---

\* Peking University

 <https://orcid.org/0000-0002-5645-3645>

 Department of Philosophy and Religious Studies, Peking University, Beijing, China.

 [ch.liu@live.com](mailto:ch.liu@live.com)

---

© The Author. Journal compilation © The Editorial Board, *Organon F*.



This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International Public License (CC BY-NC 4.0).

---

*Keywords:* Derogation; expressives; offense; pejoratives; slurs; speech acts.

## 1. Introduction

As evaluative terms, slurs are both derogatory and offensive.<sup>1</sup> On the one hand, slurs are used to derogate a group of people and members of it. For instance, calling a Chinese person ‘ch\*\*k’ seems to convey the inferior value of the Chinese. On the other hand, slurs are also used to offend people. For example, calling someone ‘ch\*\*k’ in public would outrage many who oppose racism. The goal of theories of slurs is to explain the derogatory force and the offensiveness of slurs.

However, the *offensiveness* of slurs is commonly confused with the *derogatory force* of slurs by almost all theorists. Some theories are developed to explain why slurs are “offensive” (Bolinger 2017, Jeshion 2013a, Camp 2013, Anderson and Lepore 2013b). Other theories set out to explain why slurs are “derogatory” (Croom 2011, Richard 2008, Hom 2008, Whiting 2013). Nevertheless, both varieties assume that they are explaining the same phenomenon under different terminologies. Consequently, they use “derogatory force” and “offensiveness” interchangeably. For instance, Jeshion (2013a, 244) defines the “*derogatory variation*” of utterances as the phenomenon that “utterances of different slurring terms engender different degrees of intensity of *offensiveness*.” Another example is Bolinger’s (2017, 439) definition of “*offensive autonomy*” as the phenomenon that “slurs are offensive even when the speaker does not intend the use to be *derogatory*”.<sup>2</sup> Hom’s (2012) truth-conditional content theory and Davis and McCready’s (2020) invocational content theory are the only exceptions that differentiate between being derogatory and being offensive. Nevertheless, I will argue

---

<sup>1</sup> Warning: This paper contains examples of derogatory and offensive language. All examples that mention slurs are covered up with asterisks to minimize unintended effects. I apologize for the potential offense this paper might cause.

<sup>2</sup> Another good example comes from Camp (2013, 338), whose theory of slurs claims that “they are *offensive* because their associated perspectives are negative,” but also describes slurs as “expressions that *derogate* in virtue of membership in a group like race or sex” (Camp 2013, 345).

that they either fail at explaining derogatory force or fail at explaining offensiveness.

The aim of this paper is to defend and explain the distinction between derogatory force and offensiveness. That is, being derogatory and being offensive are two different properties of slurs. My paper aims at fulfilling three goals, as follows. *First*, it will present four arguments for this distinction (§2). Derogatory force and offensiveness are different because they behave differently in non-slurs, quoted slurs, slurs in argots, and derogatory (or offensive) autonomy. *Second*, this paper will show how theories of slurs run into problems because they confuse derogatory force with offensiveness (§3). In particular, the conventional implicature theory's explanation of derogatory force does not explain why quoted slurs are offensive. Similarly, the prohibition theory's account of offensiveness can hardly explain why slurs in argots are derogatory. Even if the truth-conditional content theory and invocational content theory draw such a distinction, they nevertheless struggle with either explaining offensiveness or explaining derogatory force. *Third*, it will explain this distinction with a speech act theory of slurs (§4); slurs are derogatory and offensive because they are used to perform the illocutionary act of derogation and the perlocutionary act of offending, respectively. I will illustrate the advantage of this new explanation in avoiding the problems faced by other theories.

## 2. Distinguishing between derogatory force and offensiveness

In this section, I will present four arguments for the distinction between the derogatory force and the offensiveness of slurs. I will defend two existing arguments (§2.1 *difference in non-slurs* and §2.2 *slurs in quotations*) and offer two new arguments (§2.3 *slurs in argots* and §2.4 *offensive and derogatory autonomy*).

Before introducing the arguments, I have to clarify my usage of the term “derogatory force” and “offensiveness.” First, by “derogatory force,” I mean the property of *being derogatory*. A typical example of being derogatory is making negative remarks about someone. A word is said to possess derogatory force when it is a derogatory word. Similarly, my usage of “offensiveness” refers to the property of *being offensive*. Good examples of being

offensive include using the F-word in public or passing gas at a cocktail party. A word instantiates offensiveness when it is an offensive word. For a theory to explain slurs' derogatory force or offensiveness is for it to explain why slurs are derogatory or offensive.

### 2.1 *Difference in non-slurs*

The first argument for the distinction between derogatory force and offensiveness is that they can come apart in many expressions other than slurs. Although slurs are both derogatory and offensive, Hom and May (2013, 116) argue that many expressions can be offensive without being derogatory. Their example is a man's uttering 'You're beautiful.' to a woman passing by. Although the utterance contains only laudatory words, it is still offensive.<sup>3</sup>

A possible objection against Hom and May is that their examples are about the offensiveness of particular *utterances* (or tokens), not the offensiveness of *types* of expressions. The type of an expression is offensive when the expression *itself* is an offensive word, e.g., there is something about the F-word that makes it an offensive word. By contrast, utterances of expressions like 'You're beautiful.' offend people because the way the expressions are used, rather than the expressions *themselves*, is offensive. This objection insists that what a theory of slurs explains is the offensiveness of the *types* of slurs, i.e., what makes a word itself an offensive word. Explaining what makes utterances offensive is not the job of a theory of slurs, since any innocent word can be used to offend. Even if Hom and May's examples prove that the offensiveness of utterances is distinct from their derogatory force, their examples do not apply to the offensiveness of the *types*.

I will defend Hom and May's argument from this objection with examples of offensive but non-derogatory *types* of expressions. For instance, the British name 'Falklands' is offensive for many Argentinians because they call the disputed island 'Malvinas'. Despite its offensiveness, 'Falklands' is not a derogatory term for the island. Notice that the source of offense is the

---

<sup>3</sup> Moreover, an utterance of a slur can cease to be offensive to an oppressed racial group, once they have internalized racist ideology. Nonetheless, the slur against the oppressed group remains derogatory (Hom and May 2013, 116).

type, not the particular utterances, of the expression. For example, even a friendly utterance like ‘Falkland is a very nice island!’ may still be offensive to an Argentinian hearer. This is because he is offended by the name itself, not how it is used in an utterance.

## 2.2 *Slurs in quotations*

The second argument for the distinction is that the derogatory force of slurs behaves differently than their offensiveness in quotations. Slurs lose their derogatory force when they are mentioned (in pure or mixed quotations) but not directly used. Consider examples such as ‘‘Ch\*\*k’ is a slur for the Chinese.’ and ‘It is wrong for him to treat them like ‘ch\*\*ks.’’ Sentences like these are not derogatory against the Chinese. In pure quotations, ‘ch\*\*k’ is not used to derogate the Chinese because it is not used at all. In mixed quotations, ‘ch\*\*k’ is used to report the derogatory attitude of those who call the Chinese ‘ch\*\*ks’, rather than the speaker’s attitude.

However, slurs, as Anderson and Lepore (2013a, 36) point out, can remain offensive even if they appear in quotations. For example, Laurie Sheek, a creative writing professor at the New School in New York, mentioned the N-word in her class on *I Am Not Your N\*\*o*, a documentary on James Baldwin (McWhorter 2019). She asked the students why the title replaced ‘n\*\*er’, the word used in Baldwin’s original quote, with ‘n\*\*o’. A white student felt offended and protested her quotation of Baldwin’s use of the slur. Even if quoted slurs do not derogate a group, they can still cause offense.<sup>4</sup>

Here is a possible objection against my claim that quoted slurs are non-derogatory: it does not make sense of why quoted slurs can still be prob-

---

<sup>4</sup> There are many explanations of why quoted slurs remain offensive. For Anderson and Lepore (2013), quoted slurs are offensive because using slurs in quotations still violates prohibitions on them. By contrast, Rappaport (2020, 193) explains the offensiveness (what he calls “toxicity”) of quoted slurs in terms of their neurolinguistic effects; the phonological forms of slurs, even in quotations, can still directly trigger distinct processes in the right hemisphere like curse and taboo words. Despite the different explanations, it has been hardly disputed that quoted slurs can remain offensive.

lematic. If quoted slurs were non-derogatory, using them would be unproblematic like saying that “‘Chinese’ is the name for the Chinese.’ Nevertheless, non-bigots would be reluctant to even utter slurs in quotations, e.g., “‘Ch\*\*k’ is a slur for the Chinese.’ There are good reasons to refrain from such utterances; even quoting slurs can have bad consequences.

My response to this objection is to provide an alternative explanation of the problematic nature of quoted slurs. To say that quoted slurs are non-derogatory is not to say that they are unproblematic. Their problematic nature has another source, that is, their offensiveness. We find quoted slurs problematic because they offend many audiences, even if they do not derogate. For instance, someone who has been traumatized by incidents of being called ‘ch\*\*k’ can feel offended by merely quoting the slur. This is because hearing the slur itself often suffices to trigger negative experiences. Non-bigots would be reluctant to quote slurs because they would avoid causing such offenses and burdening the victim with traumatic experiences.

### 2.3 *Slurs in argots*

As for the third argument for the distinction, I will argue that *slurs in argots* can be derogatory without being offensive. Many slurs in argots are used to secretly derogate a particular group without causing offense. For instance, a police officer was fired for calling Boston Red Sox outfielder Carl Crawford ‘Monday’. It turns out that ‘Monday’, a seemingly innocent word, is a secret slur for African Americans in the Bostonian argot (Zimmer 2012). Similar examples include white waiters’ calling black customers ‘Canadians’, ‘cousins’, or even ‘white people’. In these cases, the point of using argots is to say something derogatory while avoiding offending people. To summarize, slurs can be derogatory without being offensive in argots, and their derogatory force is distinct from their offensiveness.

An objection against my argument is that the so-called “slurs in argots” (e.g., ‘Monday’) are not really slurs. Therefore, my examples are irrelevant and do not prove the difference between slurs’ derogatory force and offensiveness. This objection distinguishes *slurs* from non-slurs that are *used as slurs*. On the one hand, this objection insists that for an expression to be a slur, it must be a slur in a natural language like English. For instance, ‘ch\*\*k’ is a slur, and it is labeled by English dictionaries as

“derogatory.”<sup>5</sup> On the other hand, this objection holds that non-slurs can be *used as slurs*, but this does not make them slurs in a language. Suppose two speakers start to use the word ‘water’ as a slur for an oppressed group between themselves. Does this fact suddenly make ‘water’ a slur in English? Should ‘water’ be banned from public usage like ‘ch\*\*k’? The answer should be “no.” This is because being used as a slur by a few speakers does not change the meaning and the use of the word ‘water’ in English; such usage does not affect the convention of how English speakers use the word in the linguistic community. Likewise, words in argots like ‘Monday’ are used as slurs but remain non-slurs in English. After all, ‘Monday’ is not labeled as a slur or even a derogatory expression in dictionaries. One should not be accused of using slurs when one says ‘I will see you on Monday.’ In conclusion, my examples do not prove the distinction between slurs’ derogatory force and their offensiveness.

My reply to this objection is that there cannot be a clear boundary to exclude slurs in argots (e.g., ‘Monday’) from what are considered “real slurs” (e.g., ‘ch\*\*k’). Many slurs have evolved gradually from slurs in argots. For instance, ‘w\*p’, an English slur for Italians, originated from ‘guappo’ (dandy or swaggerer) in the southern Italian dialect (Zimmer 2018). It was introduced by working-class Italian immigrants to New York, where it was misheard as ‘w\*p’ by other New Yorkers. Gradually, it was adopted by other English speakers as a slur for Italians. This objection against my argument entails that ‘w\*p’ started as a non-slur in the New York workers’ argot, but it suddenly became a slur in English at a certain point. However, there was never such a magical point. This is because the process for a secret slur in an argot to become a full-blown slur in English is gradual and continuous. As the number of speakers who understand the word grows, a secret slur in the argot of a small group (e.g., New York’s immigrant workers) gradually becomes a slur in a dialect of English (e.g., the New York dialect) and eventually becomes a full-blown slur in English. This gradual process also involves a continuous growth of its offensiveness.

---

<sup>5</sup> For the definition in the Oxford English Dictionary, see ‘Ch\*\*k, n.5’. OED Online. December 2020. Oxford University Press. <https://www.oed.com/view/Entry/31779?result=5> (accessed February 17, 2021).



Slurs in argots cause no offense or less offense because very few people understand that they are derogatory. As more speakers understand these words, they become more offensive.<sup>6</sup> To summarize, there cannot be a clear boundary to exclude slurs in argots from other slurs in English.

#### 2.4 *Offensive and derogatory autonomy*

Finally, I will give the fourth argument for the distinction, i.e., that derogatory force and offensiveness differ in terms of *autonomy*. Many theories conflate “derogatory autonomy” with “offensive autonomy”, i.e., slurs are said to be derogatory or offensive regardless of the intention of the speaker (Hom 2008, Jeshion 2013a, Bolinger 2017). For instance, the “*derogatory force*” of slurs is said to be “independent of the attitudes of any of its particular speakers” (Hom 2008, 426). According to Bolinger (2017, 439), “slurs are *offensive* even when the speaker does not intend the use to be derogatory”. These theories often assume that derogatory autonomy is the same thing as offensive autonomy. A good example is the definition from Jeshion (2013a, 233): “slurs are also said to possess *derogatory autonomy*: the *offensiveness* of a use of a slurring term is ‘autonomous from the beliefs, attitudes, and intentions of individual speakers’”.

Nevertheless, I will argue that slurs exhibit offensive autonomy but *not* derogatory autonomy. That is, slurs are offensive regardless of the speaker’s intentions, but slurs’ being derogatory is affected by the speaker’s intentions. On the one hand, I acknowledge the *offensive autonomy* of slurs. Slurs can offend hearers even if the speaker does not intend to offend. For instance, saying ‘Ch\*\*k’ is a slur for the Chinese.’ with the intention to raise awareness of racism can still offend a hearer who finds hearing the slur traumatizing. On the other hand, I deny the *derogatory autonomy* of slurs. Whether using slurs is derogatory and how derogatory it is (at least sometimes) depend on the intention of the speaker. For example, the N-word can be used by African Americans to express respect, rather than derogatory attitudes, as in ‘John Brown is a straight-up n\*\*\*er.’ (Kennedy 2003, 30). Examples like this are often called “the camaraderie use” or “insular

---

<sup>6</sup> I am grateful to an anonymous reviewer for raising this point.

reclamation” (Jeshion 2020).<sup>7</sup> We do not find this utterance of the N-word derogatory because we know that the African American speakers do not intend to derogate themselves.<sup>8</sup> To summarize, offensiveness is not derogatory force because the former does not depend on the speaker’s intention like the latter.<sup>9</sup>

Here is a possible objection against my denial of derogatory autonomy: regardless of the intentions of the speakers, slurs still retain certain derogatory elements in what Jeshion (2020) calls “pride reclamation”.<sup>10</sup> For example, political activists deliberately called themselves ‘queer’. In doing so, they showed their unflinching attitude toward the derogatory uses of the bigots. According to Bianchi (2014), these “pride reclamation” uses are non-derogatory because they echo the derogatory uses of the bigots so as to show their disassociation from such derogatory uses. If no derogatory use was echoed, there would be nothing for the speaker to disassociate from.

I have two responses to this objection. First, my rejection of derogatory autonomy does not entail that intentions can always make utterances of slurs non-derogatory. Derogatory autonomy takes derogatory force to be unaffected by intentions. Therefore, it suffices to reject derogatory autonomy if there exist some cases in which utterances of slurs are made non-derogatory by intentions (e.g., the camaraderie uses from (Kennedy 2003, 30)). Second, the example of “pride reclamation” is insufficient for saving derogatory autonomy. This is because derogatory autonomy concerns whether the speaker’s *utterance* of a slur is derogatory; it is not about whether the uses of others “echoed” by the utterance are derogatory. Suppose the activists’ utterances of ‘queer’ echo the uses of the bigots. Even if

---

<sup>7</sup> Slurs can be non-derogatory in other ways. Anderson (2018) points out that the speaker can use slurs to express a neutral or positive attitude in the “referential uses” of slurs. Zeman (2021) proposes that slurs like “*ṭigan*” are non-derogatory in their “identificatory uses,” in which speakers use these slurs to identify members of their own community.

<sup>8</sup> I will explain non-derogatory utterances of slurs with my speech act theory of slurs in §4.1.

<sup>9</sup> However, there is an exception; institutional derogation does not depend on intentions. I will explain this in §4.2.

<sup>10</sup> Thanks go to Dan Zeman for pointing to this objection.

these utterances echo uses that are derogatory, the activists' utterances themselves remain non-derogatory. Here is an analogy: reporting someone's ridiculous statement does not make the report itself ridiculous. Therefore, pride reclamations do not constitute counterexamples to my rejection of derogatory autonomy.

### 3. Challenges for existing theories

In this section, I will show how conflating derogatory force and offensiveness gives rise to problems for theories of slurs such as the conventional implicature theory (§3.1) and the prohibition theory (§3.2). Although the truth-conditional content theory (§3.3) and the invocational content theory (§3.4) distinguish between being derogatory and being offensive, they still face issues such as slurs in quotations and institutional derogation.

#### *3.1 The conventional implicature theory*

According to the conventional implicature theory, slurs are derogatory because they carry derogatory conventional implicatures (Whiting 2013, Williamson 2009, Sennet and Copp 2017). For example, what is said by 'Yao is a ch\*\*k.' is the same as 'Yao is a Chinese.'; both are true if and only if Yao is Chinese. Nevertheless, using 'ch\*\*k' conventionally implicates derogatory contents such as "a noncognitive attitude of contempt (or scorn or derision or ... )" for the Chinese (Whiting 2013, 265).

However, the conventional implicature theory's account of derogatory force does not apply to offensiveness. In particular, it can hardly explain the offensiveness of slurs in quotations.<sup>11</sup> As §2.2 has shown, quoted slurs cease to be derogatory but remain offensive (Anderson and Lepore 2013a, 36). The conventional implicature theory can explain why slurs in quotations are no longer derogatory. This is because conventional implicatures are lost in quotations, e.g., "But' is a connective.'" does not conventionally

---

<sup>11</sup> Anderson and Lepore (2013a) argue against all content-based theories with quoted slurs. Although the conventional implicature theory is not specifically mentioned in their argument, their argument from quoted slurs does challenge this theory.

implicate a contrast like ‘John is rich but kind.’ Similarly, ‘‘Ch\*\*k’ is a slur for the Chinese.’ lacks the conventional implicature that the speaker has a negative attitude toward the Chinese. Unfortunately, this explanation does not apply to offensiveness. If slurs’ offensiveness originated from conventional implicatures, ‘‘Ch\*\*k’ is a slur for the Chinese.’ would be non-offensive. However, the offensiveness of slurs is not lost in quotations like conventional implicatures. There must be something other than conventional implicatures that are responsible for the offensiveness of quoted slurs.

### 3.2 *The prohibition theory*

The prohibition theory is another theory that struggles with the distinction between derogatory force and offensiveness. According to Anderson and Lepore (2013b), slurs are offensive because they are prohibited words; using slurs violates the prohibition against them. For instance, publicly calling someone ‘ch\*\*k’, which is widely forbidden, causes offense because it blatantly violates the prohibition.

However, the prohibition theory’s explanation of offensiveness does not apply to derogatory force. It faces two problems. First, it fails to account for the difference in the target of derogatory force and offensiveness. Hom (2012, 379) argues that the derogatory force of slurs targets members of a group, whereas their offensiveness targets the audience. For instance, ‘ch\*\*k’ is derogatory *against* the Chinese, but it is not merely offensive to the Chinese; it offends many non-Chinese hearers as well. Hom’s distinction presents a problem for the prohibition theory. The prohibition theory can explain why the offensiveness of ‘ch\*\*k’ does not merely target the Chinese. This is because using this slur violates the prohibition and therefore causes offense to the hearers who uphold this prohibition. Unfortunately, this is not applicable to its derogatory force; slurs are not derogatory against those who prohibit the slurs. A white hearer can be offended by someone’s usage of ‘ch\*\*k’, because he is against using racial slurs. Nevertheless, this does not make ‘ch\*\*k’ a derogatory word against white people. Therefore, there must be something other than violating prohibition that explains why derogatory force is more limited in its targets.

Second, the prohibition theory can hardly explain why slurs in argots can be derogatory without being offensive. The prohibition theory does offer

a plausible account of why slurs in argots are not offensive. This is because there is no existing prohibition on those words. In fact, the point of speaking in argots is to bypass existing prohibitions. For example, calling someone ‘Monday’ often does not cause offense because ‘Monday’ is not a prohibited word. Nevertheless, this explanation does not apply to the derogatory force. Slurs like ‘Monday’ remain derogatory despite being non-offensive and not prohibited. If it is derogatory without being prohibited, derogatory force must have sources other than prohibition.

### *3.3 The truth-conditional content theory*

Although Hom’s truth-conditional content theory offers explanations for both the offensiveness and the derogatory force of slurs, his theory nonetheless struggles with explaining offensiveness.

Hom explains slurs’ derogatory force in terms of their derogatory truth-conditional contents. For instance, ‘ch\*\*k’ means “ought to be subject to higher college admissions standards, and ought to be subject to exclusion ... , because of being slanty-eyed, and devious, and good-at-laundersing ... , all because of being Chinese” (Hom 2012, 394). Consequently, saying that ‘Yao is a ch\*\*k.’ is derogatory because it attributes derogatory properties in the content of slurs to the target. Unfortunately, it follows that sentences like ‘There are no ch\*\*ks in the building.’ cannot be derogatory because it does not attribute the negative content to anyone. This seems to entail the counterintuitive claim that nothing is wrong with such a negative existential. To address this problem, Hom distinguishes between derogatory force and offensiveness: ‘There are no ch\*\*ks in the building.’ is non-derogatory but still offensive.

Hom (2012, 402) explains the offensiveness of slurs in terms of conversational implicatures, instead of truth-conditional contents. ‘There are no ch\*\*ks in the building.’ is offensive because it carries offensive conversational implicatures. A speaker would not use ‘ch\*\*k’ unless he was committed to the existence of ‘ch\*\*ks’, i.e., people who satisfy the negative properties. Consequently, ‘There are no ch\*\*ks in the building.’ conversationally implicates that the speaker is committed to the existence of people who should be subject to discrimination because of being slanty-eyed, devious ... and being Chinese, etc.

Nevertheless, Hom's explanation of offensiveness still faces problems. First, Jeshion (2013b, 317) argues that explaining offensiveness with conversational implicature is problematic in terms of cancellability. Conversational implicatures are cancellable (Grice 1989, 39), but the offensiveness of slurs is not cancellable. For instance, saying 'There are no ch\*\*ks in the building.' does not cease to be offensive once the speaker clarifies that 'Don't get me wrong. I don't think there are people who should be discriminated against for being Chinese.'. However, if the offensiveness of 'ch\*\*k' came from conversational implicature, it would be cancellable.

Second, I will argue that Hom's explanation of offensiveness struggles with slurs in quotations. According to Hom (2012, 402), a key premise in the inferences to slurs' offensive conversational implicatures is that a speaker usually does not use a word unless he is committed to the non-empty extension of the word. However, this premise does not work for quotations. This is because speakers do not commit to the non-empty extension of a word when they quote the word. This is often why quotations can be used to deny the existence of something. For instance, 'Liberals are obsessed by the so-called 'systematic racism'.' and 'He claims that 'systematic racism' is pervasive.' does not commit the speaker to the existence of systematic racism. Quotations at most report the third party's commitment to the non-empty extension but not the speaker's commitment. Likewise, 'Ch\*\*k' is a slur for the Chinese.' and 'He says that Yao is a 'ch\*\*k'.' do not commit the speaker to the existence of 'ch\*\*ks'. Nevertheless, slurs remain offensive even in quotations. If so, their offensiveness cannot come from the kind of inference to conversational implicatures that Hom describes.

### 3.4 *The invocational content theory*

Davis and McCready's (2020) "invocational theory" takes slurs to be mixed expressives, the semantic contents of which consist of two components. First, a slur has an *at-issue semantic content*, which is simply the property of being a member of a group. Second, it has an *expressive content* (also called "invocational content"), which invokes "a complex of sociohistorical facts, attitudes, and prejudices about the group" (Davis and McCready 2020, 65). To invoke such a complex of things is to impose them to the context, i.e., updating the context to making them salient in it. Davis

and McCready's theory follows Potts's (2005, 2007) framework in treating expressivist contents as a kind of conventional implicature, which is analyzed as an extra content in addition to the at-issue semantic content.

Davis and McCready's theory also differentiates between derogatory force and offensiveness. It gives a semantic explanation of the *offensiveness* of slurs. Slurs are offensive regardless of the speaker's intention because no matter how they are used, the invoked contents of slurs are offensive by themselves.<sup>12</sup> However, Davis and McCready realize that this explanation

---

<sup>12</sup> It is unclear how Davis and McCready's theory, which treats invocational contents with Potts's framework of conventional implicatures, can explain the offensiveness of slurs in quotations. On the one hand, slurs remain offensive even if they are in quotations. However, it is not clear whether the conventional implicatures or invocational contents always project through the scope of quotations. Potts's (2005, 2007) work do not offer an analysis of conventional implicatures in quotations. Without giving justifications, Davis and McCready's (2020, 73) account of offensiveness *assumes* that conventional implicatures or invocational contents always project through quotations; "the very utterance of a slur will result in an invocation of its expressively encoded content ... no amount of embedding will help to diffuse the offensiveness that results."

However, their assumption about projection applies to mixed quotations but not pure quotations. Expressions in mixed quotations are both used and mentioned (e.g., 'John said that slurs are 'offensive'.'), whereas expressions in pure quotations are merely mentioned but not used (e.g., 'The word 'offensive' has nine letters.') (Davidson 1979). Expressions in mixed quotations can contribute their conventional implicatures to the sentence because they are still used in a way (e.g., the contrast between being rich and being kind is reported in 'John is said to be rich 'but' kind.'). Davis and McCready's assumption seems to work because they only consider slurs in mixed quotations, e.g., 'It's because he thinks of you as a S.', where S is a quoted slur (Davis and McCready 2020, 71). However, expressions in pure quotations are not used at all, and they do not seem to contribute their conventional implicatures to sentences (e.g., the conventionally implicated contrast is not reported in 'The word 'but' consists of three letters.'). It is unclear how Davis and McCready's "invocational contents," based on Potts's logic of conventional implicatures, can explain the offensiveness of pure quotations like "'Ch\*\*k' consists of five letters.' It would be surprising and ad hoc if their theory made an exception for slurs. Why would the conventional implicatures of slurs project through pure quotations, while other conventional implicatures could not?

of offensiveness does not apply to slurs' derogatory force, which to some extent depends on the speaker's intention. Consequently, their invocational content theory gives a pragmatic explanation of the *derogatory force*, i.e., "derogation is derived through reasoning about speaker attitudes" (Davis and McCready 2020, 69).

However, I argue that the invocational content theory struggles with slurs' *derogatory force*, especially in cases of institutional derogation. Davis and McCready explain non-derogatory uses of slurs in terms of the speaker's intentions; some utterances of slurs are not derogatory because hearers infer from the utterances that the speaker does not intend to endorse the offensive contents invoked by slurs. I agree with them that the speaker's intentions play a crucial role, but I do not think appealing to intentions gives the whole picture. There are cases of institutional derogation, where slurs remain derogatory despite the speaker's lack of bad intentions. Imagine that the spokesperson of a racist government calls the Chinese 'ch\*\*ks' in an official statement. Suppose we know the spokesperson is Chinese and has no intention to endorse anti-Chinese racism. Nevertheless, the slur remains derogatory against the Chinese, no matter what the speaker personally intends. This is because the spokesperson speaks on behalf of a state, not just himself. In general, slurs are not made non-derogatory by the lack of bad intentions, when they are used with institutional power (e.g., by government officials, college professors, and corporate managers).

#### 4. Explanation from a speech act theory of slurs

In this section, I will explain the distinction between derogatory force and offensiveness with a speech act theory of slurs (§4.1). I will illustrate the advantage of this explanation in avoiding the problems faced by other theories (§4.2).

##### 4.1 *The speech act theory of slurs*

*The speech act theory of slurs* explains derogatory force and offensiveness in terms of the speech acts that slurs are used to perform. It was developed in earlier works of mine to capture this basic idea: the use of slurs



is not merely to express contents but to do something against their targets (Liu 2019a, 2021).<sup>13</sup> In particular, slurs are derogatory words because they are used to perform the act of derogation, and they are offensive words because they are used to offend people. The earlier version of this theory conflated derogatory force with offensiveness like other theories. However, this paper will present a modified version, which distinguishes between derogatory force and offensiveness with different kinds of speech acts.

Slurs' being derogatory is not the same as their being offensive because to derogate is a different kind of speech act from to offend. Derogation is an *illocutionary act*, i.e., acts of doing something in saying something (e.g., promising, apologizing, asserting). By contrast, offending belongs to the category of *perlocutionary acts*, i.e., acts of producing certain effects by saying something (e.g., convincing, scaring, persuading).

The revised version of the speech act theory of slurs can be summarized as the conjunction of three theses:

The Speech Act Theory of Slurs:

- (1) *One of the uses of slurs is to derogate their targets because they are illocutionary force indicators of the illocutionary acts of derogation.*
- (2) *The other use of slurs is to offend the audiences because of their perlocutionary effects of triggering stereotypical inferences to negative properties.*
- (3) *Slurs are propositional indicators that have the same truth-conditional contributions as their neutral counterparts.*

Thesis 1 explains the *derogatory force* of slurs. Why is a slur like 'ch\*\*k' a derogatory word? According to the speech act theory of slurs, one of the two uses of slurs is to derogate their target groups. In other words, they are illocutionary force indicators of derogation. For instance, uttering 'Yao is a ch\*\*k.' is derogatory against the Chinese, because 'ch\*\*k' provides the illocutionary force of derogation against the Chinese. That is, this slur makes it explicit that 'Yao is a ch\*\*k.' should be taken as an illocutionary act of derogating the Chinese, which consists of the illocutionary force of derogation and the content, i.e., Chinese Derogation is a declarative illocutionary

---

<sup>13</sup> For more arguments for this speech act theory and its advantage over earlier speech act approaches, see (Liu 2021).

act, the point of which is to enforce a norm against the target. For instance, to derogate the Chinese by calling them ‘ch\*\*k’ is to enforce racist norms in which the Chinese deserve to be treated with violence and are deprived of their dignity, etc.

Here is a quick caveat on the function of illocutionary force indicators. An illocutionary act usually consists of an illocutionary force and a content, e.g., promising that I will come to the party has the force of promising and the content that I will come to the party. The force and the contents can come from what is called “illocutionary force indicators” (e.g., expressions like ‘I promise’, ‘I apologize’, ‘hello’ etc.) and what is called “propositional indicators” (e.g., expressions like ‘that I will come to the party’) respectively (Searle 1969, 30). Force indicators make the illocutionary force of an utterance explicit (e.g., saying that ‘I promise ...’ makes it explicit that the utterance should be taken as a promise, rather than greetings or condolences) (Searle 1996, 30). Austin (1962, 70) has an analogy of making the illocutionary force explicit: raising a hat makes it explicit that the earlier act of bowing should be taken as an act of paying respect, rather than an act of observing flowers on the ground.

Thesis 2 is added to this revised version to explain the *offensiveness* of slurs. Why is a slur like ‘ch\*\*k’ offensive? The speech act theory holds that slurs are used to perform the perlocutionary act of offending the hearers. This is because slurs produce the perlocutionary effects of triggering stereotypical inferences to negative properties. Studies find that linguistic expressions are associated with stereotypical properties that come to mind when the expressions are heard (Hare, et al. 2009, Ferretti, McRae and Hatherell 2001, Harmon-Vukić, et al. 2009). Such stereotypes trigger automatic inferences, e.g., inferring that someone is female from her being called a ‘secretary’ (Atlas and Levinson 1981). Similarly, hearing ‘ch\*\*k’ causes offense because it automatically triggers the stereotypical inferences to offensive properties associated with the Chinese, such as being slanty-eyed, devious, etc.

Thesis 3 addresses the descriptive component of slurs. As evaluative terms, slurs also make truth-conditional contributions; they are not like purely evaluative or expressive terms such as ‘damn’. In other words, slurs are propositional indicators that contribute to the truth-condition. In

particular, a slur like ‘ch\*\*k’ has the same truth-conditional contribution as its neutral counterpart, e.g., ‘Chinese’. For instance, ‘Yao is a ch\*\*k.’ is true if and only if ‘Yao is a Chinese.’ is true.

The speech act theory of slurs, by distinguishing between the two kinds of speech acts, explains the many differences between derogatory force and offensiveness introduced by the arguments in §2. First, the speech act theory can explain why derogatory force and offensiveness can come apart in non-slurs. This is because an utterance can be an illocutionary act of derogation without being a perlocutionary act of offending someone, and vice versa. Consequently, a word can be a force indicator of derogation without producing the perlocutionary effects that cause offense, and vice versa. For instance, ‘Falkland’ is an offensive word to Argentinians because it triggers stereotypical inferences to properties such as “belonging to the UK”, “military invasion”, “Argentinian defeat” etc. Nonetheless, it is not a derogatory word because it is not a force indicator used to derogate the island. Likewise, an expression can be derogatory without being offensive when it is an illocutionary force indicator of derogation without having perlocutionary effects of triggering stereotypical inferences to negative properties.

Second, my theory explains the difference between derogatory force and offensiveness in *autonomy*, i.e., why being derogatory depends on the speaker’s intention but being offensive does not. This is because the illocutionary act of derogation can be made infelicitous or unsuccessful by the speaker’s intention.<sup>14</sup> For instance, when an African American speaker utters ‘John Brown is a straight-up n\*\*er.’, this utterance is not an act of derogating African Americans. This is because successful derogation requires that the speaker must intend to impose a norm against the target. Moreover, we know the African American speaker is very unlikely to impose a racist norm against his own community in using the N-word. Unlike the derogatory force, offensiveness is autonomous or independent from the speaker’s intention. This is because whether a slur successfully produces its perlocutionary effects is determined by the hearers, not the speaker. For instance, a hearer is offended by ‘ch\*\*k’ so long as it triggers stereotypical inferences in the hearer, even if the speaker does not intend to offend.

---

<sup>14</sup> For a more detailed analysis of non-derogatory utterances of slurs, see (Liu 2019b, 2021). For the exception of institutional derogation, see the following section.

In addition, my theory also explains why slurs in quotations remain offensive and why slurs in argots can be non-offensive. I will give detailed explanations in the following section since these two issues challenge other theories.

#### 4.2 *Advantages over other theories*

The advantage of my explanation of derogatory force and offensiveness is that it avoids the problems faced by other theories.

In comparison to the *conventional implicature theory*, the speech act theory of slurs does a better job of explaining why quoted slurs, despite being non-derogatory, can be offensive. Slurs in quotations are non-derogatory because illocutionary force indicators in quotations cannot be used to perform speech acts (e.g., saying that ‘I promise’ is an illocutionary force indicator.’ does not make a promise). Likewise, ‘Ch\*\*k’ is a slur for the Chinese.’ does not have the use to derogate the Chinese. However, quoted slurs remain offensive because quoted words can still trigger stereotypical inference in the hearers. Upon hearing ‘Ch\*\*k’ is a slur for the Chinese.’, the slur automatically triggers stereotypical inferences to negative properties such as being slanty-eyed and devious.

As for the *prohibition theory*, the speech act theory enjoys two advantages over it. First, my theory explains why the derogatory force and the offensiveness of slurs have different targets. For instance, ‘ch\*\*k’ is derogatory against the Chinese because it is an illocutionary force indicator of derogation against the Chinese. However, ‘ch\*\*k’ is not merely offensive to the Chinese. This is because slurs also trigger stereotypical inferences to negative properties in non-Chinese hearers. Second, my theory performs better at explaining the offensiveness of slurs in argots. Slurs like ‘Monday’ can be derogatory without being offensive. This is because it is an illocutionary force indicator of derogation against the African Americans in the Bostonian argot. Nevertheless, it is not offensive because it does not trigger stereotypical inferences in the hearers. For the hearers who are not familiar with the argot, ‘Monday’ is not associated with the stereotypical properties of African Americans.

Although Hom’s *truth-conditional content theory* also differentiates between derogatory force and offensiveness, my theory performs better than

it in two ways. First, it explains the non-cancellability of slurs' offensiveness. One cannot explicitly cancel the offensiveness of 'There are no ch\*\*ks the building.' with 'Don't get me wrong. I don't think there are people who should be discriminated against for being Chinese.' This is because offensiveness arises from the perlocutionary effects of stereotypical inferences. Such inferences are automatic and involuntary. Therefore, it cannot be canceled by opting out of the cooperative principle. Second, my theory has no problem in explaining the offensiveness of quoted slurs, as I have just described.

Finally, my theory differs from Davis and McCready's invocational content theory in two aspects. To be illocutionary force indicators is not the same as expressing "invocational contents". First, these two theories disagree over the effects of the slurs. For the invocational content theory, the primary effects of the invocational content are *linguistic*; their job is to change the contexts, e.g., making certain contents salient. Contextual changes affect the interpretations of context-sensitive expressions and the directions of discourses. By contrast, the speech act theory does not take the primary effects of slurs to be merely linguistic; their use is to perform the acts of derogation that change the *normative status* of the target (e.g., rights, obligations, what are allowed and forbidden to do), not just how we talk about them. For instance, the slur 'ch\*\*k' is harmful because it enforces anti-Chinese norms that deny the rights of the Chinese, license violence against them, and deprive them of dignity.

Second, these two theories have different views on the sources of derogatory force. The invocational content theory explains the derogatory force of slurs in terms of inferences to the speaker's intentions. It follows that slurs are not derogatory when we infer that the speaker has no bad intentions. While the speech act theory allows such inferences to affect the derogatory force of slurs, it does not take these inferences to be the source of derogatory force. Instead, slurs are derogatory because they are used to perform the speech act of derogation, which can be made unsuccessful by the lack of derogatory intentions in certain cases.

By limiting the role of the speaker's intention, the speech act theory of slurs is immune to the problem of institutional derogation faced by the invocational content theory. Like many other illocutionary acts (e.g.,

personal apologies and official apologies), derogation can be distinguished into *personal derogation* (e.g., making negative remarks to my friends at a party) and *institutional derogation* (e.g., issuing an official condemnation as the president). Personal derogation does not require institutional power to perform, but its success requires that the speaker intends to impose a norm against the target. By contrast, the successful performance of institutional derogation has requirements on the role of the speaker (e.g., the speaker must occupy offices such as being the president or the spokesperson to derogate on behalf of an institution), but not her personal intentions. In conclusion, my speech act theory holds that slurs, when used with institutional power, can be derogatory regardless of the speaker's intention. This is because the successful performance of *institutional derogation*, unlike personal derogation, does not depend on intentions.

## 5. Conclusion

In this paper, I have achieved three goals. First, I have presented four arguments for the distinction between slurs' derogatory force and offensiveness. For instance, being derogatory is not the same as being offensive because slurs in argots, e.g., 'Monday', are used to derogate their targets without causing offense. Second, I have argued that failing to draw such a distinction has given rise to problems for theories of slurs. The prohibition theory, for example, cannot explain the derogatory force in terms of the prohibition. This is because slurs in argots (e.g., 'Monday') are used to derogate their targets even if there is no prohibition on them. Third, I have offered a new explanation of this distinction with a speech act theory of slurs, which distinguishes between the two uses of slurs, i.e., performing the illocutionary act of derogation and performing the perlocutionary act of offending. The advantage of this new explanation is that it avoids the difficulties of other theories, such as slurs in quotations and slurs in argots.

These considerations, I hope, justify the distinction between slurs' derogatory force and offensiveness. There is more than one way for slurs, as well as other terms, to be expressive.

## References

- Anderson, Luvell. 2018. "Calling, Addressing, and Appropriation." In *Bad Words*, edited by David Sosa, 6–26. Oxford: Oxford University Press.  
<https://doi.org/10.1093/oso/9780198758655.001.0001>
- Anderson, Luvell, and Ernie Lepore. 2013a. "Slurring Words." *Noûs* 47 (1): 25–48.  
<https://doi.org/10.1111/j.1468-0068.2010.00820.x>
- Anderson, Luvell, and Ernie Lepore. 2013b. "What Did You Call Me? Slurs as Prohibited Words." *Analytic Philosophy* 54 (3): 350–63.  
<https://doi.org/10.1111/phib.12023>
- Atlas, Jay David, and Stephen C. Levinson. 1981. "It-clefts, Informativeness and Logical Form: Radical Pragmatics (Revised Standard Version)." In *Radical Pragmatics*, edited by Peter Cole, 1–62. New York: Academic Press.
- Austin, J. L. 1962. *How to Do Things with Words*. Clarendon Press.
- Bianchi, Claudia. 2014. "Slurs and Appropriation: An Echoic Account." *Journal of Pragmatics* 66: 35–44. <https://doi.org/10.1016/j.pragma.2014.02.009>
- Bolinger, Renée Jorgensen. 2017. "The Pragmatics of Slurs." *Noûs* 51 (3): 439–62.  
<https://doi.org/10.1111/nous.12090>
- Camp, Elisabeth. 2013. "Slurring Perspectives." *Analytic Philosophy* 54 (3): 330–49. <https://doi.org/10.1111/phib.12022>
- Croom, Adam M. 2011. "Slurs." *Language Sciences* 33 (3): 343–58.  
<https://doi.org/10.1016/j.langsci.2010.11.005>
- Davidson, Donald. 1979. "Quotation." In *Inquiries into Truth and Interpretation*, 79–92. Oxford: Oxford University Press.
- Davis, Christopher, and Elin McCready. 2020. "The Instability of Slurs." *Grazer Philosophische Studien* 97 (1): 63–85. <https://doi.org/10.1163/18756735-09701005>
- Ferretti, Todd R., Ken McRae, and Andrea Hatherell. 2001. "Integrating Verbs, Situation Schemas, and Thematic Role Concepts." *Journal of Memory and Language* 44 (4): 516–47. <https://doi.org/10.1006/jmla.2000.2728>
- Grice, Paul. 1989. *Studies in the Way of Words*. Harvard University Press.
- Hare, Mary, Michael Jones, Caroline Thomson, Sarah Kelly, and Ken McRae. 2009. "Activating Event Knowledge." *Cognition* 111 (2): 151–67.  
<https://doi.org/10.1016/j.cognition.2009.01.009>
- Harmon-Vukić, Mary, Sabine Guéraud, Karla A. Lassonde, and Edward J. O'Brien. 2009. "The Activation and Instantiation of Instrumental Inferences." *Discourse Processes* 46 (5): 467–90.  
<https://doi.org/10.1080/01638530902959661>
- Hom, Christopher. 2012. "A Puzzle About Pejoratives." *Philosophical Studies* 159 (3): 383–405. <https://doi.org/10.1007/s11098-011-9749-7>

- Hom, Christopher. 2008. "The Semantics of Racial Epithets." *Journal of Philosophy* 105 (8): 416–40. <https://doi.org/10.5840/jphil2008105834>
- Hom, Christopher, and Robert May. 2013. "Moral and Semantic Innocence." *Analytic Philosophy* 54 (3): 293–313. <https://doi.org/10.1111/phib.12020>
- Jeshion, Robin. 2020. "Pride and Prejudiced." *Grazer Philosophische Studien* 97 (1): 106–37. <https://doi.org/10.1163/18756735-09701007>
- Jeshion, Robin. 2013a. "Expressivism and the Offensiveness of Slurs." *Philosophical Perspectives* 27 (1): 231–59. <https://doi.org/10.1111/phpe.12027>
- Jeshion, Robin. 2013b. "Slurs and Stereotypes." *Analytic Philosophy* 54 (3): 314–29. <https://doi.org/10.1111/phib.12021>
- Kennedy, Randall. 2003. *Nigger: The Strange Career of a Troublesome Word*. New York: Vintage Books.
- Liu, Chang. 2021. "Slurs as Illocutionary Force Indicators." *Philosophia* 49 (3): 1051–65. <https://doi.org/10.1007/s11406-020-00289-0>
- Liu, Chang. 2019a. *Derogatory Words and Speech Acts: An Illocutionary Force Indicator Theory of Slurs*. PhD dissertation: The University of Western Ontario.
- Liu, Chang. 2019b. "Slurs and the Type-Token Distinction of Their Derogatory Force." *Rivista Italiana di Filosofia del Linguaggio* 13 (2): 63–72. <https://doi.org/10.4396/12201902>
- McWhorter, John. 2019. "The Idea That Whites Can't Refer to the N-Word." *The Atlantic*. August 27. Accessed November 12, 2020. <https://www.theatlantic.com/ideas/archive/2019/08/whites-refer-to-the-n-word/596872/>
- Potts, Christopher. 2007. "The Expressive Dimension." *Theoretical Linguistics* 33 (2): 165–98. <https://doi.org/10.1515/TL.2007.011>
- Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.
- Rappaport, Jesse. 2020. "Slurs and Toxicity." *Grazer Philosophische Studien* 97 (1): 177–202. <https://doi.org/10.1163/18756735-09701010>
- Richard, Mark. 2008. *When Truth Gives Out*. Oxford: Oxford University Press.
- Searle, John R. 1996. "What Is a Speech Act?" In *The Philosophy of Language* (third edition), edited by A. P. Martinich, 130–40. Oxford University Press.
- Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.
- Sennet, Adam, and David Copp. 2017. "Pejoratives and Ways of Thinking." *Analytic Philosophy* 58 (3): 248–271. <https://doi.org/10.1111/phib.12100>
- Whiting, Daniel. 2013. "It's Not What You Said, It's the Way You Said It: Slurs and Conventional Implicatures." *Analytic Philosophy* 54 (3): 364–77. <https://doi.org/10.1111/phib.12024>



- Williamson, Timothy. 2009. "Reference, Inference, and the Semantics of Pejoratives." In *The Philosophy of David Kaplan*, edited by Joseph Almog, and Paolo Leonardi, 137–59. Oxford University Press.
- Zeman, Dan. 2021. "A Rich-lexicon Theory of Slurs and Their Uses." *Inquiry*. <https://doi.org/10.1080/0020174X.2021.1903552>
- Zimmer, Ben. 2018. "'Wop' Doesn't Mean What Andrew Cuomo Thinks It Means." *The Atlantic*. April 23, 2018. Accessed November 9, 2020. <https://www.theatlantic.com/politics/archive/2018/04/wop-doesnt-mean-what-andrew-cuomo-thinks-it-means/558659/>
- Zimmer, Ben. 2012. "How Did 'Monday' Become a Racist Slur?" *Boston Globe*. July 29, 2012. Accessed November 3, 2020. <https://www.bostonglobe.com/ideas/2012/07/28/how-did-monday-become-racist-slur-how-did-monday-become-racist-slur/Mf4fQEVcXabGKHFaDMZ4NO/story.html>

## Rethinking Slurs: A Case Against Neutral Counterparts and the Introduction of Referential Flexibility

Alice Damirjian\*


Received: 4 December 2020 / Revised: April 14 2021 / Accepted: 2 June 2021

*Abstract:* Slurs are pejorative expressions that derogate individuals or groups on the basis of their gender, race, nationality, religion, sexual orientation and so forth. In the constantly growing literature on slurs, it has become customary to appeal to so-called “neutral counterparts” for explaining the extension and truth-conditional content of slurring terms. More precisely, it is commonly assumed that every slur shares its extension and literal content with a non-evaluative counterpart term. I think this assumption is unwarranted and, in this paper, I shall present two arguments against it. (i) A careful comparison of slurs with complex or thick group-referencing pejoratives lacking neutral counterparts shows that these are in fact very hard to distinguish. (ii) Slurs lack the referential stability of their alleged neutral counterparts, which suggests that they are not coreferential. Developing (ii) will involve introducing a new concept which I regard as essential for understanding how slurs behave in natural language: referential flexibility. I shall support my claims by looking at historical and current ways in which slurs and other pejorative terms are used, and I shall argue that both etymological data and new empirical data support the conclusion that the assumption of neutral counterparts not only is unwarranted but obscures our understanding of what slurs are, and what speakers do with them.

---

\* Stockholm University

 <https://orcid.org/0000-0003-4410-2087>

 Universitetsvägen 10D, 114 18 Stockholm, Sweden.

 [alice.damirjian@philosophy.su.se](mailto:alice.damirjian@philosophy.su.se)

---

© The Author. Journal compilation © The Editorial Board, *Organon F*.



This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International Public License (CC BY-NC 4.0).

---

*Keywords:* Neutral counterparts, pejorative language, philosophy of language, semantics, slurs.

## 1. Introduction

In the constantly growing literature on slurs, it has become customary to appeal to so-called “neutral counterparts” for explaining the extension and truth-conditional content of slurring terms. More precisely, it is commonly assumed that every slur shares its extension and literal content with a non-evaluative counterpart term. Paradigmatic examples of slurs with respective neutral counterparts in the literature are: ‘n\*gg\*r’ with ‘Black person’, ‘k\*ke’ with ‘Jew’ and ‘ch\*nk’ with ‘Chinese’.<sup>1</sup>

This assumption often takes the form of an identity thesis, which states that slurs are extensionally and truth-conditionally equivalent with their relevant neutral counterparts. In his often-cited paper, Whiting summarizes his view in the following way: “[...] what is said by the use of a slur is what is said by the use of its neutral counterpart, whereas what is derogatory about the use of a slur is the claim or, as I prefer, attitude conventionally implicated by it” (2013, 376). In a similar vein, Vallée argues that: “If S is an ethnic slur in language L, then there is a non-derogatory expression G in L such that G and S have the same extension” (2014, 79). It is this suggested equivalence between slurs and their alleged neutral counterparts that will be addressed in this paper.

Seeing as this assumption about the truth-conditional content of slurring terms has become so widely accepted, the debate surrounding slurs has largely come to be about what *other* properties slurs should be ascribed, that can account for their offensive nature. Some take the derogatory content of a slur to be expressed by way of a conventional implicature, so that it becomes part of the slur’s conventional meaning, but not its truth-

---

<sup>1</sup> CONTENT WARNING: Here, I have chosen to transcribe the slurs with asterisks in order to avoid causing unnecessary harm, and I will continue to do so throughout this paper. However, for clarity, less common slurs have been left unedited, and I will also not transcribe slurs occurring in quotes from other authors or sentence examples taken from social media.

conditional content (cf. Whiting 2013; Williamson 2009). Others prefer to analyze it as a presupposition (cf. Schlenker 2007; Cepollaro 2015), or as a violation of prohibition (Anderson and Lepore 2013a, 2013b). Yet others suggest that we must focus on what speakers are doing when they *choose* to use a slur instead of its non-derogatory counterpart term. Examples of such accounts are Camp's (2013, 2018) perspectivist account and pragmatic accounts like Bolinger's (2017) and Nunberg's (2018).<sup>2</sup> To illustrate, Ashwell has described Camp as giving us

[...] the most persuasive reason for thinking that there has to be a neutral alternative term for a word to be a slur – that it is only in contrast with an alternative neutral term that the bigot can signal their bigoted perspective. (Ashwell 2016, 238)

Of course, how much a specific account relies on the existence of neutral counterparts varies. Prohibitionist and conventional implicature views are, for example, not necessarily reliant on neutral counterparts.<sup>3</sup> However, many accounts require the existence of the neutral counterpart for their preferred analysis to work, and this is especially pressing for accounts that place weight on the *optionality* of slurs, e.g. Bolinger's and Nunberg's.<sup>4</sup> But if the assumption of neutral counterparts turns out to be non-viable, all proponents of accounts that rely on them to some degree will have to re-think how they explain the truth-conditional content of slurring terms.

In this paper, I shall argue that the pervasive idea that neutral counterparts can play this semantic role in our theories of slurs is unmotivated and that it obscures our understanding of what slurs are and what speakers do with them. This will be supported by a body of evidence that casts serious doubt upon the assumption that the extension of a slur is identical to the extension of some salient neutral counterpart. Then, if we are not warranted in assuming that every slur is extensionally equivalent with a neutral counterpart, we cannot assume that every slur shares its literal meaning with

---

<sup>2</sup> Undeniably, this list is very far from complete.

<sup>3</sup> Since it in principle could be argued that a slur is prohibited, or has the relevant conventional implicature, without it being extensionally or truth-conditionally equivalent with a neutral counterpart.

<sup>4</sup> See Falbo (2021) for a more detailed explanation for why this is.

one either. The central claim is thus that the identity thesis is deeply flawed, and the important consequence of this is that we must rethink how we analyze the semantics of slurs, so as to be able to account for and predict the data that will be provided here. The paper will not be proposing an alternative positive account of the meaning of slurs; instead, the aim is to provide insights needed for moving forward in the field.

## 2. Neutral counterpart theories

Diaz Legaspe (2018) distinguishes between two ways in which theories of slurs rely on neutral counterparts. Firstly, we can identify a weaker position, widely accepted in most of the literature, which Diaz Legaspe calls the *Application Neutral Counterpart Thesis* (AT). Secondly, there is a stronger position not as widely accepted, called the *Referential Neutral Counterpart Thesis* (RT). The two theses are defined as follows (Diaz Legaspe 2018, 235):

AT: For every slurring expression  $e$  there is a neutral counterpart  $N\text{Ce}$  such that  $N\text{Ce}$ 's correct application criteria are identical to  $e$ 's correct application criteria.<sup>5</sup>

RT: For every slurring expression  $e$  there is a neutral counterpart  $N\text{Ce}$  such that the class of individuals referred to by  $N\text{Ce}$  (call it  $\{N\text{Ce}\}$ ) is identical to the class of individuals referred to by  $e$ .<sup>6</sup>

AT is intended to describe accounts similar to Hom's (2010, 2012) and Hom and May's (2013), which pack the derogatory content of slurs into their

---

<sup>5</sup> Diaz Legaspe argues that the  $N\text{Ce}$  can be an actual or potential neutral counterpart to  $e$ , "potential" meaning that the neutral counterpart might not have been found yet or that it is for some reason unavailable. This, however, strikes me as strange. Either there already exists a neutral counterpart to  $e$  in the language or  $e$  lacks a neutral counterpart, we should not just be able to state that  $e$  could be given a neutral counterpart.

<sup>6</sup> The notation might strike the reader as odd. To clarify,  $N\text{Ce}$  is not a set, but an abbreviation for the "neutral counterpart to  $e$ ."  $\{N\text{Ce}\}$  is denoting a set: the class of individuals referred to by  $N\text{Ce}$ . With that said, I will continue to use these notations as introduced by Diaz Legaspe.

semantics. On Hom's account, every slur targets some specific group picked out by its neutral counterpart, but the slur and the neutral counterpart are still taken to diverge in meaning and extension (in fact, the extension of a slur is assumed to be the empty set). Hence, AT is silent regarding the truth-conditional contribution of a slur. "Correct application" is therefore not intended to be understood in a strong sense.<sup>7</sup> Instead, "correct application" simply makes it a necessary condition that the individual referred to by *e* is a member of {N*Ce*} for *e* to have been correctly applied. So, the thesis does predict that it is a linguistic mistake to use *e* for someone outside of {N*Ce*}, but it does not claim *e* and N*Ce* to be intersubstitutable *salva veritate*. In contrast, RT states that the extension of a slur will be the same as that of its neutral counterpart, and this is supposed to be a consequence of the terms' making the same truth-conditional contribution.<sup>8</sup>

RT has turned out to be a common approach for explaining the semantics of slurring terms, and this is standardly the case for the views included here in §1. Consequently, it is primarily the identity thesis at the heart of RT which this paper will target. Even so, since AT does predict that it is a linguistic mistake to use *e* for an individual outside of {N*Ce*}, AT will also be sensitive to the arguments presented here.

### 3. Neutral counterpart skepticism

There have been attempts to show that the whole idea of co-referentiality between a slur and a non-pejorative synonym is misguided. For example, Croom (2015) has argued that empirical data shows that slurs cannot be coreferential with a neutral counterpart. Croom provides the reader with

---

<sup>7</sup> For a strong reading would suggest co-referentiality. AT is intended to describe and provide "a conception of correct application that does not amount to truth-conditional identity" (Diaz Legaspe 2018, footnote 5, 135).

<sup>8</sup> This is not explicitly stated in RT but it is assumed that co-referentiality amounts to truth-conditional equivalence. "In turn, RT predicts that for every slur and every associated neutral counterpart, both will contribute the same set of individuals to the truth-conditions of the utterance in which they occur" (Diaz Legaspe 2018, 235).

four examples, taken from academic research, literature and comedy, of when speakers are employing slurs in ways inconsistent with what RT predicts.<sup>9</sup> These examples are, even if Croom does not call them that, examples of *referential restriction* (a phenomenon we will come back to in §4.2); cases of slurring in which the slur is used not to pick out all of {NCe} but only a subset. For example, when ‘n\*gg\*r’ is applied to some *but not all* Black people, or when ‘f\*ggot’ is deemed not to apply to all gay men, but only to *some* male homosexuals. This leads Croom to the conclusion that

[...] the fact that the slur *faggot* is differentially used so that it is often applied to *some but not all* male homosexuals suggests that the slur *faggot* and the descriptor *male homosexual* are in fact *not coreferential expressions with precisely the same extension* at all. (2015, 32)

If slurs are not coreferential with a neutral counterpart, then a slur cannot be said to share its meaning with one either, Croom argues, and any theory of slurs will have to respect that fact. This is an observation that will be of importance throughout this paper.

Croom is not the only one to criticize the identity thesis at issue here. Ashwell (2016) has argued that all accounts of slurs that assume neutral counterparts will fail to generalize over all terms we are inclined to regard as slurs. According to Ashwell this is best shown when considering gendered slurs, such as ‘slut’ or ‘bitch’, which have not been given the same attention as racial or ethnic slurs in the literature.

Some might want to argue (e.g., Diaz Legaspe 2018) that ‘bitch’ and similar gendered slurs have ‘woman’ as their neutral counterpart, but it appears unlikely that bigoted individuals would hold that “*all* women are bitches” in the same way as they would hold that “*all* Jews are k\*kes” or “*all* Blacks are n\*gg\*rs.” Of course, ‘bitch’ and ‘slut’ are generally applied to women, but one is not a ‘slut’ for being a woman. One is a ‘slut’ in virtue of something else, Ashwell claims (2016, 234), something that has to do with sex. However, “woman who has sex with a lot of partners” cannot function as a neutral counterpart, it is not free of pejorative association nor

---

<sup>9</sup> The examples will not be presented here but can be found in (Croom 2015, 32-34).

is it neutral because “a lot of partners” is dependent on what is assumed to be the appropriate number of sexual partners (Ashwell 2016, 234-35).

Are gendered slurs actually slurs then, or some other kind of pejorative? “They are not slurs” is the answer scholars like Nunberg (2018) would give: they are what is called *hybrid words* and therefore not slurs in the strict sense (see §4.1 in this paper). Objections like Ashwell’s can therefore be dismissed by claiming that the terms brought forward as counterexamples are not in fact the kind of pejoratives that we are interested in; not slurs but some other type of pejorative. However, if one wants to go down that road one must be careful, as DiFranco (2015) has pointed out:

They [those assuming neutral counterparts] should not simply insist that the class of conventional slurring words is, by definition, restricted to words and phrases whose truth-conditional content is identical to that of their neutral counterparts. Doing so would beg the question by illicitly presupposing NC [truth-conditional equivalence]. (2015, 33)

One should not simply claim that the class of slurs is exactly that class of non-complex slurring terms that share their truth-conditional content with a neutral counterpart. Such a claim would have to be motivated. It could be motivated if: (i) It can be shown that the class of slurs is significantly large and clearly distinguished from other kinds of pejoratives lacking neutral counterparts; or (ii) we have strong evidence that slurs express the same thing, on a truth-conditional level, as their alleged neutral counterparts. For the remainder of this paper, I shall argue that no such motivation exists.

#### 4. Against neutral counterparts

In the following sections I will argue for two points, each of which, and especially in combination, should be regarded as severely undermining the credibility of appealing to the identity thesis in one’s theory of slurs. The first point has to do with the characterization of slurs, and I will argue that we cannot distinguish a distinct class of slurs from other pejoratives lacking neutral counterparts. The second point is that slurs can be used to refer in



flexible ways, a feature of slurs that their alleged neutral counterparts lack. As argued by Croom, speakers can use slurs with referential restriction to refer to a subset of the set {N*C*e}, but I will argue that speakers also use slurs to refer outside of {N*C*e}, in ways that cannot be disregarded as linguistic mistakes. I will call this phenomenon *referential expansion*. This will lead us to the conclusion that slurs are used, and can be used, in flexible ways, both restrictedly and expansively. I will call this feature *referential flexibility*.<sup>10</sup>

Referential flexibility can be observed both historically and in present-day use, and I will therefore spend a fair amount of time on etymological examples and speaker examples that I have found on the internet. Already, I would like to make clear that the point of the etymological examples is not say that slurring terms somehow have kept the meaning they have had historically. Rather, it is to show that their application criteria have been, and still are, flexible – or at least not non-flexible in the sense RT and AT suggest.

#### 4.1 *Distinguishing the non-distinct distinctions*

*Hybrid words* and *umbrella derogatives* are terms usually regarded as slurs, but which appear to lack neutral counterparts.<sup>11</sup> Hybrid words, as

---

<sup>10</sup> It should be noted that once referential flexibility has been introduced, talk of referential expansion will soon prove problematic, for such talk implies the existence of a neutral counterpart that can be expanded upon. The reader should therefore keep in mind that the characterization of referential expansion as it is introduced here is not intended to serve any explanatory purposes apart from the argumentative role it plays within the scope of this paper. Once we have dropped neutral counterparts, there is no further need to discuss the kind of referential expansion I introduce. Hopefully, all of this will become clear after §4.2.

<sup>11</sup> The discussion here will primarily focus on umbrella derogatives, as they pose the greatest challenge to theories relying on neutral counterparts. In principle, hybrid words could be dealt with by holding that the additional evaluative content they incorporate does not contribute to determining the terms' extensions. This way out is suggested by Jeshion (2013, 234-235). However, for scholars like Nunberg (2018) who hold that hybrid words should not be conflated with slurs, the discussion here about hybrid words is highly relevant.

defined by Nunberg (2018), are terms that do not only categorize but mix categorization and attitude, so that some strong evaluation of the referent is part of the content of the term that cannot be found in any non-evaluative synonym. They are sometimes referred to as “thick terms”, precisely because they incorporate an evaluation or a stereotypical content within their semantics. Examples are ‘wetback’ for (illegal) Mexican migrants, ‘JAP’ (‘Jewish American Princess’) for spoiled Jewish women, and ‘Uncle Tom’ for Black people who behave in subservient ways towards white people. ‘Slut’, ‘bitch’ and most other disparaging terms for women are also taken to fall into this category (see Nunberg 2018, 249-250).<sup>12</sup>

Umbrella derogatives are pejoratives for collections of distinct groups which become problematically grouped together, to the extent that the terms can have no neutral counterpart.<sup>13</sup> Clear-cut examples are ‘slope’, mostly used for East Asians, ‘wog’ standardly used for any non-white foreigner, ‘dago’ sweepingly used for Italians, Spaniards and Portuguese, and ‘gook’ for foreigners, especially those of East Asian descent.<sup>14</sup> There rarely exists a synonym for referring to the same group, the group itself is hard to distinguish, and even if we were to stipulate a potential neutral counterpart it would prove difficult to find a non-evaluative one.<sup>15</sup> For that reason,

---

<sup>12</sup> Whether or not we are dealing with something similar to thick ethical concepts, such as ‘brave’ and ‘generous’, can be debated, but nothing in this discussion hinges on whether such a similarity exists. For the purposes of this paper, it is enough to observe that there is a structural similarity insofar as both hybrid words and standard examples of thick terms exhibit this sort of hybrid nature. I thank the anonymous referee who drew my attention to this unclarity.

<sup>13</sup> Note that this does not entail that umbrella derogatives lack semantic meaning, just that there does not exist any salient neutral counterpart with which they could be said to share literal meaning.

<sup>14</sup> Jeshion (2016, 135) also notes that slurs like ‘wop’, ‘dago’ and ‘gook’ fail to possess neutral counterparts.

<sup>15</sup> Why cannot, for example, ‘non-white foreigner’ play the neutral counterpart-role? The problem is not that this description does not pick out a category that it could be said to refer to, the problem is that ‘wog’ does not mean ‘non-white foreigner’. When I say that ‘wog’ is used for non-white foreigners it is a simplified account of how speakers tend to use the term, not that it is coreferential with ‘non-white foreigner’. Furthermore, ‘non-white foreigner’ provides us with a large collection of

a theory of slurs that places weight on neutral counterparts in its analysis will fail to deal with umbrella derogatives.

According to Nunberg (2018), these terms should not be confused with slurs, which in a stricter sense belong to the class of pejoratives that have a non-slurring counterpart. This is important for his account to work, and for many others. However, is it reasonable to think that such a class of slurs can be said to exist in any well-defined sense, such that slurs, as a well-considered category, can be singled out as Nunberg suggests?

When looking closer, we can observe interesting similarities between umbrella derogatives like ‘gook’ and pejoratives like ‘ch\*nk’ and ‘j\*p’, both perceived as unproblematic examples of slurs. ‘Gook’ stands out in that the term has come to play many different xenophobic roles in a relatively short period of time. As a pejorative, it has not only been used against people of East Asian descent but for any foreigner (from an American perspective) and any foreign language (i.e., not English).<sup>16</sup> In his encyclopedia of swearing, Hughes explains that “[...] its semantic history combines hostility toward outsiders with great flexibility in application” (2006, 207). But ‘gook’ is not the only term that has played varying roles; ‘ch\*nk’ does in fact have a similar, but not as striking, history of flexibility in application.

During the 1849 California Gold Rush a great number of Chinese immigrants arrived in America to work as indentured laborers, and the resentment towards the immigration of cheap labor within the group of white native-born laborers resulted in several names for the out-group. The dominant derogative for the Chinese immigrants was ‘ch\*nk’ (Hughes 2006, 75-

---

distinct groups that should not be conflated and so arbitrarily referred to as a single group. Thus, even if we stipulated that ‘non-white foreigner’ gives us the meaning of ‘wog’, it would not be neutral in the sense required. Imagine a speaker uttering ‘There were so many k\*kes in the park today.’ The (supposedly) equivalent utterance ‘There were so many Jews in the park today’ can reasonably be said to be neutral in the sense required (if there is nothing in the context of utterance to suggest otherwise). Arguably, the same is not the case for ‘There were so many non-white foreigners in the park today.’ Standardly, this is taken to be what distinguishes slurs from umbrella derogatives.

<sup>16</sup> The etymological data presented here, and in other places in the paper, comes from Green’s *Chambers Slang Dictionary* (2008) and Hughes’ *An Encyclopedia of Swearing* (2006).

76). The term was used for Chinese people but also for people with Chinese features, and thus more generally for any person of East Asian descent. It was especially from 1942 that ‘ch\*nk’ began to be applied to any East Asian person, which coincides with the evolution of ‘j\*p’. The abbreviation ‘j\*p’ for ‘Japanese’ was common from around the 1850s, but not necessarily offensive and not exclusively used for Japanese. But after Pearl Harbor, terms of abuse arose rapidly and ‘j\*p’ was simultaneously used for Japanese people and as a slang for being sneaky or a bad surprise. Other people with similar appearances were conflated too, others from the Far East, all of whom were labeled ‘gooks’ (Hughes 2006, 262).

So, are these slurs better described as umbrella derogatives rather than actual slurs in the relevant sense? The answer depends on how speakers today use the slurs, which is an empirical question, but it is plausible to assume that slurs are used in this loose way quite often. Consider this testimony from a man of Cambodian descent:

This guy in DC just skipped me in line at 7/11 [7-Eleven] and then proceeded to call me a ch\*nk – multiple times. I told him he had skipped me in line and that’s when he got aggressive. Anti-Asian racism is real and it’s fucked up. I’m okay, just a jarring experience.<sup>17</sup>

The example illustrates two important points: (i) That the speaker deems it relevant to utilize the term ‘ch\*nk’, not knowing or not caring about the actual nationality of the targeted person, and (ii) that the targeted person himself makes a point of saying that the speaker’s words were anti-Asian, rather than anti-Chinese.

Additionally, one could also argue that ‘j\*p’ functions more like a hybrid word, i.e., that one can observe it possessing the properties deemed distinctive of a hybrid word. Nunberg claims that hybrid words are distinguished from slurs by the fact that the evaluative content present in hybrid words can be contested, but not restated without a feeling of redundancy. In contrast to actual slurs, hybrid words carry their evaluative content within

---

<sup>17</sup> <https://tinyurl.com/tjsb5mc>, accessed 13 April 2020. All URLs have been shortened as they are sometimes very long, but will redirect the reader to where the examples can be found.

their conventional meaning, and thus a sense of redundancy will arise in sentences like:

- (1) Uncle Toms are really obsequious.
- (2) Bitches are malicious women.

Sentences like (1) and (2) should rightly elicit the reaction “So what else is new?” Nunberg argues (2018, 249); it is already part of the meaning of calling somebody an ‘Uncle Tom’ that they are obsequious or subservient. Yet, with actual slurs that feeling of redundancy does not occur, hence sentences (3) and (4) should appear informative, and indeed common.

- (3) N\*gg\*rs are so lazy.
- (4) I don’t like that k\*ke, he’s very greedy.

Primarily, this is one of Nunberg’s arguments against conventional implicature views, for if the derogatory content is part of the conventional meaning of slurs, then this redundancy should be present in any construction similar to (1)–(4), and not only those involving hybrid words. But it also provides us with a clear description of hybrid words.

Still, it does not seem all that easy to weed out the hybrids from the slurs. Turning back to ‘j\*p’, the rapid expansion of the use of ‘j\*p’ during the Second World War as, simultaneously, a word for the Japanese soldiers and the Japanese living in the United States (and indeed anyone East Asian looking), and as a slang term for sneaky things and bad surprises, might suggest that the term could be treated as a hybrid word. There is reason to think that the current use of ‘j\*p’ carries that evaluative content within its meaning, conventionally, so that one should react with “So what else is new?” in response to (5).

- (5) The j\*ps are so sneaky.

This is not necessarily to say that ‘j\*p’ is a hybrid word. The point is to illustrate that the distinction between slurs, hybrid words and umbrella derogatives is not crystal clear – in fact, it is not clear at all.

Depending on how you approach the terms, you might get different intuitions about how they are used. If you choose to only study the cases in which ‘ch\*nk’ is used for Chinese people, trying to figure out what the term means, you will probably feel that ‘ch\*nk’ refers only to Chinese people.

But if you instead start to look at all the cases in which ‘ch\*nk’ and ‘j\*p’ are used more broadly, you will probably agree that these cases attest to the terms functioning more like umbrella derogatives or hybrid words.

#### 4.2 Arguments for referential flexibility

Diaz Legaspe (2018) attempts to tackle the problematic aspects of referential restriction, which were illustrated in Croom’s argument above, and tries to defend both AT and RT against it. To repeat: referential restriction occurs when a slur is used to refer to a subset of {N<sub>Ce</sub>}. A commonly used example of the phenomenon is the statement (6) made by comedian Chris Rock.

(6) I love Black people, but I hate niggers.<sup>18</sup>

Referential restriction poses a problem for theories relying on RT, Diaz Legaspe argues, because if slurs always share their truth-conditional content with their associated neutral counterparts, then sentences like (6) should be contradictory. However, sentences like (6) are common and appear informative, so some slurs do seem to be able to refer to a narrower class than {N<sub>Ce</sub>}. Further, Diaz Legaspe observes, some slurs even appear to always refer to a sub-class of their neutral counterpart, such as gendered slurs.

To try and solve this, Diaz Legaspe proposes modifications to AT and RT in order to restore the link between slurs and their neutral counterparts. Diaz Legaspe’s proposal is, roughly, that there can be particular contexts in which a slur *e* (e.g. a racial slur), whose reference is {N<sub>Ce</sub>}, can be used to refer to a subset of {N<sub>Ce</sub>}, but there are also some slurs (e.g. all gendered slurs) which *always* refer to a subset of the class picked out by their neutral counterparts.<sup>19</sup> Thus, racial slurs and gendered slurs should be understood differently, but even so, gendered slurs can be assigned neutral counterparts (holding that the slurs just always refer to a subset of the set picked out by

<sup>18</sup> The example appears in (Diaz Legaspe 2018, 236 and 243). This is also one of Croom’s examples (2015, 33) and it appears in: (Nunberg 2018, footnote 12, 247), (Rappaport 2019, 810), (Jeshion 2013, 233 and 238-239), (Anderson and Lepore 2013b, footnote 3, 43), to name a few occurrences.

<sup>19</sup> This approach is partly aimed at solving the problems with gendered slurs set up by Ashwell (2016).

their neutral counterparts). These are the modified versions of the two theses, which are assumed to hold whether you accept RT and AT or just AT, and for all types of slurs (Diaz Legaspe 2018, 248-249):

*Negative AT*: For any slur  $e$  there is a  $\text{NCe}$  such that every member of  $\{\text{NCe}\}$  can be correctly called an ‘ $e$ ’.

Furthermore, *only* members of  $\{\text{NCe}\}$  can correctly be called an ‘ $e$ ’, with the exception of metaphorical uses (Diaz Legaspe 2018, 248).<sup>20</sup> To call someone outside of  $\{\text{NCe}\}$  an ‘ $e$ ’, will count as a linguistic mistake. The second condition is:

*Positive AT\**: For every  $e$  there is a  $\text{NCe}$  such that every member of  $\{\text{NCe}\}$  could potentially be called an ‘ $e$ ’.<sup>21</sup>

That is, all in  $\{\text{NCe}\}$  can potentially be called an ‘ $e$ ’, without it amounting to a linguistic mistake. Seeing as some slurs, like gendered slurs, always refer to a subset of  $\{\text{NCe}\}$  “normal” RT will not hold for them, Diaz Legaspe argues, but if one generally wants RT to hold for slurs one can appeal to *Restricted RT*:

*Restricted RT*: Whenever “ $o$  is an  $e$ ” is true, “ $o$  is a  $\text{NCe}$ ” is also true.

When we are dealing with gendered slurs, however,  $e$  and  $\text{NCe}$  will not be interchangeable in the other direction.

Now, if a case of referential restriction is a case in which a slur  $e$  is used to refer to a subset of  $\{\text{NCe}\}$ , as in sentence (6), then referential expansion is a case in which  $e$  is used to refer outside of  $\{\text{NCe}\}$ , in a way that cannot be disregarded as a linguistic mistake. Of course, such expansions are not allowed once one has accepted the identity thesis, and thus cases of refer-

<sup>20</sup> After having introduced the condition, Diaz Legaspe writes “[...] only women can be correctly called ‘sluts’ [...]” (2018, 248).

<sup>21</sup> This condition (Positive AT\*) is a modified version of Positive AT:

Positive AT: For every  $e$  there is a  $\text{NCe}$  such that every member of  $\{\text{NCe}\}$  can be called an ‘ $e$ ’.

A condition which Diaz Legaspe claims not to hold, since it is assumed that some slurs always refer to a subset of  $\{\text{NCe}\}$ . But, she concludes, potentially everyone in  $\{\text{NCe}\}$  could be called ‘ $e$ ’.

ential expansion should rightly be treated as counterexamples to the identity thesis. To be clear, the referential expansion that interests us here does not occur from referencing outside of {NCe} in any intentionally incorrect or figurative way by alluding to stereotypes. Neither is referential expansion in our sense simply some linguistic effect a speaker can evoke by referencing outside of {NCe}.

Slurs can expand and gain flexibility in different ways. One way, which is discussed in Hughes' encyclopedia, is when a slur develops from its basic noun function to be used as an adjective or a verb. Standardly, this is seen as an indicator of the slur having become very assimilated into the language (Hughes 2006, 149). However, this is not the kind of expansion that interests us either.<sup>22</sup> What interests us is when slurs are actually used to refer to individuals outside of the set {NCe}. The example which supported treating 'ch\*nk' as an umbrella derogative is an example of referential expansion.

A category of slurring terms not generally discussed are derogatory terms for disabled people, which are interesting because, like gendered slurs, they too seem difficult to capture within the frameworks of preexisting theories. Such slurs also exhibit the phenomenon of referential expansion because they are commonly used expansively in several different ways. Consider for example the slurring term 'spastic', and alterations such as 'spaz' and 'spazzie', and how they have come to be used. 'Spastic' was first used as a non-derogatory term for people with cerebral palsy, subject to muscle spasm or spasticity, but also became a pejorative for that same group. Understanding 'spastic' as a slur, we should be able to find a salient neutral counterpart in the language. Arguably, its neutral counterpart should be, if

---

<sup>22</sup> The possibility of such morphological transformations could however be seen as supporting my claims. Consider a case in which the noun 'k\*ke' is transformed into a verb so that sentences like "He kiked his way to the job" become possible. Or a case of someone uttering "He pulled a jap," or perhaps (a phrase common on the internet) "He's so spazzy." Arguably, what the speaker is intending to convey in cases like the above has little to do with predicating group membership. Rather some descriptive value is intended, and this might suggest that some descriptive, evaluative content is indeed part of the slurs' conventional meaning, also when they are used as nouns. However, it is too complicated a matter for me to be able, within the scope of this paper, to say anything more about it than that it might support my claims.



we were to assign one, ‘people with cerebral palsy’ or ‘people subject to spastic paralysis.’ However, this is not solely how the term tends to be utilized.

Two types of extended uses can be observed, one is an example of what Jeshion calls *G-extending* uses (2013, 238); when speakers use ‘spastic’ of someone *they believe not to have* any physical disability but whom the speaker *wants* to ascribe stereotypical properties associated with people with cerebral palsy, such as being jumpy, clumsy, incapable or incompetent. This kind of expanded use does not necessarily pose a problem for theories postulating neutral counterparts, and therefore it is not this kind of use that interests us.<sup>23</sup> The second kind is when the slur is used for people with other disabilities (people that do not have cerebral palsy), i.e. more broadly of any person with a disability, to derogate them in virtue of that. This can include people with similar symptoms, such as seizures, or other diagnoses associated with similar behaviors, e.g. people living with Tourette syndrome or ADHD. In such cases, we have much less reason to suspect that the relevant speaker is intending to say something that is literally incorrect for some type of linguistic effect. The more reasonable explanation for why this is possible is that the extension of ‘spastic’ is flexible to such a degree that it is not restricted to people with cerebral palsy. That explains why speakers can use ‘spastic’ in this broad way. To exemplify, responding to a post discussing how to calm people with ADHD and overactive children, a Reddit user writes:

(7) Lol just read a book u spastic [sic].<sup>24</sup>

Diaz Legaspe’s (2018) modified versions of AT and RT were shaped to be compatible with, and even explain, the existence of referential restriction but they are not compatible with referential expansion (nor referential flexibility). If we assume that slurs have the same neutral counterparts as they have been ascribed in previous literature, then every case of referential expansion – every case in which a speaker refers outside of {NCe} with e –

<sup>23</sup> It does not seem problematic to state that this kind of use involves an intentionally incorrect statement, and therefore it does not call into question the appropriateness of the neutral counterpart assumed.

<sup>24</sup> <https://tinyurl.com/vfasc19>, accessed 13 April 2020.

should amount to linguistic mistakes following *Negative AT*. *Restricted RT* will also not hold, because if an utterance of “o is an e” is a case of referential expansion it will not also be true that “o is a NCe.” Only *Positive AT\** (or even the unmodified *Positive AT*) will still hold, for even if individuals outside of {NCe} can be called an ‘e’, it would still be true that all in {NCe} can potentially be called an ‘e’, but in isolation this condition is very weak. That would open the possibility for many sets to play the neutral counterpart-role, and we would have no way of determining between them, and if it is underdetermined between a number of sets, and potentially all of them could play the role, then referential flexibility is what we get.

Moreover, these kinds of observations cannot be restricted to specific types of slurs, as some might want to claim. Even ‘n\*gg\*r’ shows an interesting history of flexibility in application. From about the 17th century ‘n\*gge\*r’ was used to refer to Black people, but primarily slaves. It then evolved to be used for any non-white, around the 19th century, and even more generally for any foreigner, and it also began being used for the Aborigines in Australia.<sup>25</sup> Expanded uses, outside of the alleged {NCe} for ‘n\*gg\*r’, that is ‘Black person’ or ‘African American’, thus seem to have been present for a long time. This tendency is also what allows for expansions with additional content prevalent today, such as ‘sand n\*gg\*r’ or ‘dune n\*gg\*r’ for people of Middle Eastern descent, ‘curry n\*gg\*r’ for Indians, as well as ‘bush n\*gg\*r’ used for Native Americans, Africans and Aborigines. These are terms that might seem important for the speakers to distinguish, and please excuse the phrase, *just what kind of ‘n\*gg\*r’ it is they are referring to*.<sup>26</sup>

<sup>25</sup> See (Green 2008, 914-15), for a more detailed account of how the meaning of ‘n\*gg\*r’ has changed over time.

<sup>26</sup> On a side note, there is an interesting question here for compositional semantics about the function of the modifiers ‘sand’, ‘dune’, ‘red’ etc. What are the adjectives actually doing? Normally, an adjective functions so to restrict the extension of the noun it modifies, so that when ‘ball’ is modified with ‘red’ to form ‘red ball’ one has restricted the denotation of ‘ball’ to only include those balls that are red. That is however not what is happening here, if one maintains that ‘n\*gg\*r’ denotes Black people then ‘red n\*gg\*r’ cannot be said to restrict the noun (given that ‘red n\*gg\*r’ is not, standardly, used for Black people at all). Is the modifier then completely changing the extension of the noun? Does it take us outside of the extension of

This can be exemplified with an interaction described by a man working at a hotel in Canada, who describes himself as originally from India. Another man, upset by what he regards as bad customer service, proceeds to call the man ‘sand n\*gg\*r’ and threatens him as in (8).

- (8) You fucking sand nigger, do you want me to call my boys and have a picnic at your hotel, you fucking piece of shit.<sup>27</sup>

When asked to leave he continues:

- (9) You fucking nigger, go back to your country, you asshole.<sup>28</sup>

Sentences (8) and (9) not only illustrate that the speaker in this situation deems it relevant to use ‘n\*gg\*r’ and ‘sand n\*gg\*r’ against an Indian man, but also that he regards ‘sand n\*gg\*r’ as interchangeable with ‘n\*gg\*r’.

In that sense ‘n\*gg\*r’ is similar to ‘coon’. In American and British English ‘coon’ has generally been used as a derogatory term for Black people, but not exclusively. Rather, ‘coon’ is similar to ‘wog’ since it has been used more generally for any person of color. It is therefore not surprising that we can see a tendency among speakers to use ‘dune coon’ interchangeably with ‘sand n\*gg\*r’ and ‘dune n\*gg\*r’ to refer to Middle Eastern people.

In another Reddit post, a person contemplates the racial slurs that others have called them, which mainly have been words targeting Middle Easterners, such as ‘sand n\*gg\*r’ and ‘osama’, as well as ‘curry n\*gg\*r’. Then, in the same post, they observe that:

- (10) Seriously, you can totally change the direction of the n-word just by tacking on a certain word in front of it.<sup>29</sup>

---

‘n\*gg\*r’, in similar ways as ‘fake’ could be said to do when used as a modifier? That might be a possible approach for those assuming neutral counterparts, but it would appear ad hoc to claim that ‘red’ has the capacity to modify ‘ball’ and ‘n\*gg\*r’ in two very different ways. But, if we allow for the meaning of ‘n\*gg\*r’ to be broad and flexible like an umbrella derogative’s, we might be able to keep some version of the normal understanding of what the adjectives do.

<sup>27</sup> <https://tinyurl.com/ukdvewz>, accessed 13 April 2020.

<sup>28</sup> <https://tinyurl.com/ukdvewz>, accessed 13 April 2020.

<sup>29</sup> <https://tinyurl.com/vtccllx>, accessed 13 April 2020.

These uses of ‘n\*gg\*r’ are not G-extending uses nor are they changing the term itself, but they help change the term’s *direction*, that is, its extension. They specify, as stated above, *just what kind of ‘n\*gg\*r’ it is they are referring to* – which is completely in line with understanding slurs like ‘n\*gg\*r’ as referentially flexible.

Now, some might want to object and say that my arguments against neutral counterparts will render slurs radically flexible. It was, for instance, pointed out to me by an anonymous referee that referential flexibility will render slurs extensionally unlimited. I very much agree that this is a worry. Of course, there must be some constraints on how slurs can be applied; ethnic slurs derogate groups on the basis of their ethnicity, and it would not seem right (linguistically) to call someone a ‘n\*gg\*r’ because of their sexual orientation. Even so, understanding that slurs are flexible, to such a degree that the extension of ‘n\*gg\*r’ is not restricted to any well-defined group, will help explain why they are so difficult to account for philosophically. Referential flexibility is not the end of the story, it invites us to re-think how the meaning of slurs can be accounted for, and my conviction is that when we have found ways of reconceptualizing the meaning of slurs, we will be in a better position to explain why slurs are offensive.

In a recent paper, Falbo (2021) has argued that we must be cautious about assuming that neutral counterparts can play any fundamental or systematic role in explaining the offensiveness of slurs, for in a range of examples neutral counterparts seem unable to do the job they were supposed to do. This observation is right, but we must also accept the stronger claim that neutral counterparts seem unable to play any fundamental role in explaining the truth-conditional content of slurs. Moving forward, we must begin by understanding that racial slurs are used by racists, and *what racists do* is to group people together in arbitrary and insensitive ways – and slurs are a medium for doing just that.

## 5. Conclusion

To answer the questions leading up to §4: (i) Is the class of slurs significantly large and clearly distinguished from other kinds of pejoratives lacking neutral counterparts? The answer is *no*, many of our most discussed

slurs, such as ‘ch\*nk’ and ‘n\*gg\*r’, are in fact hard to distinguish from umbrella derogatives, and others, like ‘j\*p’, are not sufficiently different from hybrid words. Even if we could find some slurs that escape the arguments presented here, they would be rare and therefore any theory of slurs that is only capable of accounting for those few will be insufficiently general. (ii) Do we have strong evidence that slurs express the same thing, on a truth-conditional level, as their alleged neutral counterparts? The answer to this question is also *no*. Since we have evidence for referential flexibility both AT and RT are in trouble, because if slurs and neutral counterparts have non-identical extensions, then they cannot be truth-conditionally equivalent. The conclusion of this paper is thus that the assumption of neutral counterparts is problematic, and that the identity thesis lacks relevant motivations.

### Acknowledgments

I would like to thank Dan Zeman and the reviewers for their encouragement and constructive comments, my supervisors Anders Schoubye and Kathrin Glüer-Pagin for their support, and Jesper Olsson for his positivity and sharp eyes.

### References

- Anderson, Luvell, and Ernie Lepore. 2013a. “What Did You Call Me? Slurs as Prohibited Words.” *Analytic Philosophy* 54 (3): 350–63.  
<https://doi.org/10.1111/phib.12023>
- Anderson, Luvell, and Ernie Lepore. 2013b. “Slurring Words.” *Noûs* 47 (1): 25–48.  
<https://doi.org/10.1111/j.1468-0068.2010.00820.x>
- Ashwell, Lauren. 2016. “Gendered Slurs.” *Social Theory and Practice* 42 (2): 228–39. <https://doi.org/10.5840/soctheorpract201642213>
- Bolinger, Renée. 2017. “The Pragmatics of Slurs.” *Noûs* 51 (3): 439–62.  
<https://doi.org/10.1111/nous.12090>
- Camp, Elisabeth. 2013. “Slurring Perspectives.” *Analytic Philosophy* 54 (3): 330–49. <https://doi.org/10.1111/phib.12022>
- Camp, Elisabeth. 2018. “A Dual Act Analysis of Slurs.” In *Bad Words: Philosophical Perspectives on Slurs* edited by David Sosa, 29–59. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198758655.001.0001>

- Cepollaro, Bianca. 2015. "In Defence of a Presuppositional Account of Slurs." *Language Sciences* (52): 36–45. <https://doi.org/10.1016/j.langsci.2014.11.004>
- Croom, Adam M. 2015. "The Semantics of Slurs: A Refutation of Coreferentialism." *Ampersand* (2): 30–8. <https://doi.org/10.1016/j.amper.2015.01.001>
- Diaz Legaspe, Justina. 2018. "Normalizing Slurs and Out-Group Slurs: The Case of Referential Restriction." *Analytic Philosophy* 59 (2): 234–55. <https://doi.org/10.1111/phib.12129>
- DiFranco, Ralph. 2015. "Do Racists Speak Truly? On the Truth-Conditional Content of Slurs." *Thought: A Journal of Philosophy* 4 (1): 28–37. <https://doi.org/10.1002/tht3.154>
- Falbo, Arianna. (2021). "Slurs, Neutral Counterparts, and What You Could Have Said." *Analytic Philosophy* (not yet included in an issue). <https://doi.org/10.1111/phib.12217>
- Green, Jonathon. 2008. *Chambers Slang Dictionary*. Edinburgh: Chambers Harrap Publishers Ltd.
- Hom, Christopher. 2010. "Pejoratives." *Philosophy Compass* 5 (2): 164–85. <https://doi.org/10.1111/j.1747-9991.2009.00274.x>
- Hom, Christopher. 2012. "A Puzzle About Pejoratives." *Philosophical Studies* 159 (3): 383–405. <https://doi.org/10.1007/s11098-011-9749-7>
- Hom, Christopher, and Robert May. 2013. "Moral and Semantic Innocence." *Analytic Philosophy* 54 (3): 293–313. <https://doi.org/10.1111/phib.12020>
- Hughes, Geoffrey. 2006. *An Encyclopedia of Swearing: The Social History of Oaths, Profanity, Foul Language, and Ethnic Slurs in the English-Speaking World*. New York: M. E. Sharpe.
- Jeshion, Robin. 2013. "Expressivism and the Offensiveness of Slurs." *Philosophical Perspectives* 27 (1): 231–59. <https://doi.org/10.1111/phpe.12027>
- Jeshion, Robin. 2016. "Slur Creation, Bigotry Formation: The Power of Expressivism." *Phenomenology and Mind* (11): 130–9. [https://doi.org/10.13128/Phe\\_Mi-20113](https://doi.org/10.13128/Phe_Mi-20113)
- Nunberg, Geoff. 2018. "The Social Life of Slurs". In *New Work on Speech Acts* edited by Daniel Fogal, Daniel Harris, and Matt Moss, 237–95. Oxford: Oxford University Press. <http://doi.org/10.1093/oso/9780198738831.001.0001>
- Rappaport, Jesse. 2019. "Communicating with Slurs." *The Philosophical Quarterly* 69 (277): 795–816. <https://doi.org/10.1093/pq/pqz022>
- Schlenker, Philippe. 2007. "Expressive Presuppositions." *Theoretical Linguistics* 33 (2): 237–45. <https://doi.org/10.1515/TL.2007.017>
- Vallée, Richard. 2014. "Slurring and Common Knowledge of Ordinary Language." *Journal of Pragmatics* (61): 78–90. <http://dx.doi.org/10.1016/j.pragma.2013.11.013>

- 
- Whiting, Daniel. 2013. "It's Not What You Said, It's the Way You Said It: Slurs and Conventional Implicatures." *Analytic Philosophy* 54 (3): 364–77. <https://onlinelibrary.wiley.com/doi/10.1111/phib.12024>
- Williamson, Timothy. 2009. "Reference, Inference, and the Semantics of Pejoratives." In *The Philosophy of David Kaplan* edited by Joseph Almog, and Paolo Leonardi, 137–58. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195367881.001.0001>

## The Moral Status of the Reclamation of Slurs

Bianca Cepollaro\*

Received: 18 December 2020 / Accepted: 4 June 2021

*Abstract:* While prototypical uses of slurs express contempt for targets, some reclaimed uses are associated with positive evaluations. This practice may raise concerns. I anticipate this criticism in what I dub the Warrant Argument (WA) and then defend the legitimacy of this kind of reclamation. For the WA, standard pejorative uses of slurs are problematic for assuming unwarranted connections between descriptive properties (e.g., being gay) and value judgements (e.g., being worthy of contempt). When reclaimed uses of slurs express a positive evaluation of their targets—the WA goes—reclamation fails to challenge the unwarranted link between descriptive properties and value judgements, and merely reverses the evaluation polarity from negative to positive. So, the WA concludes, reclaimed uses of slurs evaluating targets positively for belonging to a certain group make a similar moral error as derogatory uses of slurs (sections 2-3). The WA could lead us to condemn reclamation. To resist this conclusion, I draw a parallel with affirmative action, arguing that it can be morally permissible to balance an existing form of injustice by temporarily introducing a countervailing mechanism that prima facie seems to violate the norm of equality: even if the WA were right, it wouldn't constitute an argument against the moral permissibility of reclamation in the case of most slurs (section 4). This line of

---

\* Vita-Salute San Raffaele University

 <https://orcid.org/0000-0003-2695-2125>

 Università Vita-Salute San Raffaele Via Olgettina, 58, 20132 Milano, Italy.

 [cepollaro.biancamaria@univr.it](mailto:cepollaro.biancamaria@univr.it)





argument in defense of pride reclamation may also serve to debunk the myths of reverse racism and reverse sexism (section 5).

*Keywords:* Hate speech; polarity reversal; reclamation; reverse racism; reverse sexism; slurs.

## 1. Introduction

While prototypical uses of slurs often express contempt for targets, some reclaimed uses of such epithets are associated with a positive evaluation of the target class. This practice—let’s call it ‘pride reclamation’—may raise concerns under certain readings. In this paper, I anticipate this criticism—summarized in what I dub the Warrant Argument (WA)—and defend the legitimacy of this kind of reclamation. According to the WA, what is problematic with standard pejorative uses of slurs is that they assume an unwarranted connection between descriptive properties (such as being gay, being Italian, being Jewish and so on) and value judgements (being bad, being worthy of contempt and the like). When reclaimed uses of slurs express pride and/or a positive evaluation of their targets—the WA goes—they fail to challenge such a wrong link between descriptive properties and value judgements, and merely manage to reverse the polarity of the value judgement from negative to positive, from contempt to pride. So, the WA concludes, reclaimed uses of slurs expressing a positive evaluation of the slur’s targets make a similar moral error as ordinary derogatory uses of slurs (sections 2-3). The WA could lead us to condemn or even ban pride reclamation, just like we do with derogatory uses of slurs. To resist this conclusion, I draw a parallel with the mechanisms of affirmative action, to argue that it can be morally permissible to balance an existing form of injustice by temporarily introducing a countervailing mechanism that *prima facie* seems a violation of the norm of equality: the analogy with affirmative action suggests that, even if the WA were right, it wouldn’t constitute an argument against the moral permissibility of pride reclamation in the case of most slurs (section 4). Finally, I show how this line of argument in defense of pride reclamation may also serve to debunk the myths of reverse racism and reverse sexism (section 5).

## 2. The moral mistake of slurring

Slurring is to derogate people on the basis of their belonging to a certain category that typically has to do with race, sexual orientation, gender, nationality, religion and so on (see i.e. Saka 2007; Hom 2008; Richard 2008). What distinguishes slurs from other pejoratives such as ‘asshole’ or ‘jerk’ is that only the former target people qua members of a social group. A slur like ‘wop’ does not merely derogate an Italian person, but attacks her because she is Italian.

Scholars have developed different theories as to how to analyze slurs and their pejorative content (see Hess forthcoming for a rich but concise survey). In this paper, I stay neutral with respect to two central aspects: (i) how slurs convey the derogatory content with which they are associated (i.e., whether the derogatory content is lexically encoded or not; and if it is, how), and (ii) how such a derogatory content should be spelt out. With respect to these issues, I will just adopt two general assumptions that are compatible with most existing accounts of slurs: (i) standard uses of epithets are systematically associated with derogatory contents and (ii) the derogatory content of slurs—however it is conveyed—amounts to something roughly like ‘bad for being P’, where ‘P’ is a descriptive property such as being gay, being Black, being Italian, etc. Let me briefly elaborate on both points.

Regarding (i), the use of slurs conveys pejorative contents across contexts, regardless of the intentions of the speaker; moreover, it conveys such contents even when slurs are embedded under semantic operators. In fact, the derogatory content of epithets displays a striking feature that has attracted a good deal of attention from linguists and philosophers of language in the past years, namely projection (see i.a. Potts 2005; Schlenker 2007; Anderson and Lepore 2013; Jeshion 2013a, 2013b; Bolinger 2015; Nunberg 2018; Camp 2018). ‘Projection’ refers to the fact that the derogatory content of slurs survives in the interaction with many semantic operators, i.e., projects out of semantic embedding. Take the sentence ‘Lucia is a wop’. The derogatory content towards Italian people—however we prefer to spell it out—survives when we embed the sentence under negation, in the antecedent of a conditional, in a question or a modal: ‘Lucia is not a wop’, ‘If Lucia is a wop, her brother is too’, ‘Is Lucia a wop?’, ‘Lucia might be a

wop'. As one can see, the derogatory content of slurs is very hard to suspend, at least in their non-reclaimed uses.

As for (ii), it suggests that what slurs do in general is to express a negative evaluation (often characterized as contempt) based on some descriptive properties (e.g., being gay, being Black, being Italian, etc.) that do not warrant per se a negative evaluation. Using slurs thus involves a morally wrong connection between a descriptive property and a value judgement. Such a link between the descriptive property and the evaluation is not merely correlational: from the perspective that the use of slurs encourages, targets do not just happen to be bad; they are bad because they satisfy a certain descriptive property (e.g., being Italian, gay, Black and so on). Suppose that research finds out that every single Italian person is bad in some respects: if you know that someone is Italian, you automatically know that s/he is bad in some way. This correlational claim that every Italian is bad in some respect would still differ from the racist assumption conveyed by uses of 'wop' that an Italian is bad for being Italian. In this paper I assume that the idea that being from a certain country (as well as having a certain sexual orientation, gender, religion, etc.) makes an individual worthy of a negative evaluation is morally wrong.

The morally problematic assumption that the use of slurs expresses and promotes, together with the endurance and pervasiveness of their derogatory content, makes epithets particularly toxic and dangerous. It is not surprising that slurs are typically banned from the public debate in liberal democracies (see i.a. Stanley 2015) and some scholars advocated silentist positions, according to which any occurrence of slurs—including mere mention—is offensive and should be therefore banned (Anderson and Lepore 2013). There are non-derogatory occurrences of slurs, though, that usually survive most kinds of censorship: they are the so-called reclaimed uses of epithets, to which we now turn.

### **3. Pride reclamation of slurs and the Warrant Argument criticism**

Reclamation is the phenomenon for which speakers, typically members of the target group, can use a slur in such a way that the pejorative is not

offensive anymore in those contexts; in contrast, reclaimed slurs often convey positive contents and attitudes about the target class, they typically express pride, intimacy, solidarity, and camaraderie (see Tirrell, 1999; Brontsema, 2004; Croom 2011, 2013, 2014; Bianchi 2014; Miščević and Perhat 2016; Ritchie 2017; Anderson 2018; cf. Burnett 2020 and Zeman 2021).

Even though it is unclear that all instances of reclamation do so, at least some reclaimed uses of derogatory epithets seem to switch the standard evaluative content of slurs from negative into positive, thus turning contempt into pride (i.a., Kennedy 2002; Ritchie 2017; Jeshion 2020, Popa-Wyatt 2020). Under a very general characterization, these reclaimed uses of slurs convey something along the line of ‘good for being P’ (Ritchie 2017). I dub these uses ‘pride reclamation’.

Many scholars got interested in reclamation because it constitutes an instance of meaning change (a temporary or stable change, according to different views) that challenges most existing accounts of slurs. Others, like Bianchi (2014), underline further interesting aspects: reclamation has proved so far one of the most effective tools to get rid of the toxic and harmful powers of slurs. For Bianchi (2014), while reclamation weakens the toxicity of slurs, silentism risks to worsen it: this is why reclamation should be encouraged and reclaimed uses of slurs should not be banned nor censored. In this picture, reclamation is presented as a possible way to counteract racism, homophobia and discrimination, and scholars have pointed out the benefits of reclamation with respect to empowerment (see i.a. Croom 2014). This view is also supported by empirical studies (see for instance Galinsky et al. 2003; Galinsky et al. 2013), according to which self-labeling—applying a slur to oneself—has important empowering effects.

Not everyone agrees, however: scholars such as Bailey et al. (1998) have illustrated a wide range of attitudes vis-à-vis the practice of reclamation, from very positive to very negative. The detractors of reclamation have various strings to their bow: they can argue that reclamation is self-defeating because it ultimately disguises self-contempt (Kennedy 2002). They may worry that it ends up legitimizing the use of slurs in a dangerous way (Herbert 2015), for instance by suggesting that certain terms are harmless and can be used inconsiderately. It could be maintained that reclamation

contributes to the linguistic segregation of targets, thus worsening their overall social segregation, and so on.

In this paper, I anticipate and reject a different worry, summarized in what I call the Warrant Argument (WA). According to the WA, instances of reclamation conveying a positive evaluation of the target class ('good for being P') are morally problematic in that they end up making a similar moral mistake to the one standard pejorative uses of slurs make. Recall what we have observed in section 2, namely that slurs express negative judgements of a subject on the basis of a descriptive property that does not warrant a negative evaluation per se. Those who find this morally problematic might also agree that being from a certain country (as well as having a certain sexual orientation, gender, religion, etc.) does not by itself make anyone good, just as much as it doesn't make one bad. If one acknowledges that being Italian (or gay or Black, etc.) does not warrant a negative value judgement—the WA goes—they should also acknowledge that it does not warrant a positive evaluation either. The WA does not assume that because a descriptive property P does not warrant a negative evaluation, then it cannot warrant a positive one at the same time—this would be trivially false. Rather, it takes it that the particular descriptive properties picked out by prototypical slurs (in relation to race, sexual orientation, gender, etc.) do not warrant per se any value judgement at all (neither positive nor negative). Since, as we saw, standard uses of slurs convey a negative evaluation and the reclaimed uses (at least some of them) convey a positive evaluation, the Warrant Argument—applied to 'wop'—goes as follows:

1. Standard uses of a slur like 'wop' convey the idea that being Italian in itself warrants a negative evaluation, and this is wrong, because being Italian never justifies any value judgement: there is nothing bad or good per se in being Italian.
2. Certain reclaimed uses of 'wop' convey the idea that being Italian is per se good, thus that it warrants a positive evaluation. This is wrong, because being Italian never justifies any value judgement: there is nothing bad or good in itself in being Italian.
3. So, certain reclaimed uses of 'wop' that convey a positive evaluation of Italians qua Italians make a similar moral mistake to the one made by standard negative uses of slurs.

The WA leads one to conclude that both derogatory and reclaimed slurs encapsulate a moral mistake. This in turn may prompt the conclusion that also reclaimed slurs should be banned. Indeed, most existing accounts of reclamation suggest that what reclamation does to slurs (especially at its early stages) is to reverse the polarity of the evaluation (or expressive content, depending on the account one favors) from negative to positive. On these accounts, reclamation (at least at its earlier stages) does not challenge the unwarranted link—associated with slurring—between a property such as being Italian, being Black, being gay, etc. and a value judgement.

One could stop the WA right here by suggesting that there is an important difference between pride and contempt, between expressing positive and negative evaluations. One could say that the WA shouldn't lead to conclude that reclamation has to be banned simply because while violence and harm follow from hate speech, nothing similar is likely to follow from reclamation. In fact, pride per se doesn't need to involve a feeling of superiority over everyone else (cf. recognition vs. appraisal respect in Darwall 1977). This strategy to stop the WA is more problematic than it seems, because it's not so clear that positively connoted social terms are harmless: think, for instance, of how 'Aryan' was used by Nazis. Think what it would be like if slurs against white people were reclaimed (?). Invoking a deep asymmetry between social terms associated with positive rather than negative contents may not be enough to stop the WA from banning reclamation. In this paper, I signal a longer route to do so. The advantage of this longer journey is that by granting more to the WA, its conclusion in favor of the moral permissibility of reclamation should be appealing to a larger audience. Finding answers to the WA that can be shared widely is especially important, given the connection with the myth of reverse racism or sexism to which I'll turn in section 5.

This said, if we accept, with the WA, that certain kinds of reclamation encapsulate and endorse a faulty connection between descriptive properties regarding social groups and unwarranted evaluations, could such a practice of reclamation be nevertheless morally permissible? In the next session I present a parallel with affirmative action, aimed at maintaining that reclamation is morally permissible, notwithstanding the criticism from the Warrant Argument.

Before moving to the next section, let me remark that not all accounts of reclamation need to deal with the WA, as not all approaches grant that reclaimed uses of slurs involve ‘polarity reversal’, as Jeshion (2020) calls it. While many proposals (e.g., Ritchie 2017, Jeshion 2020, Popa-Wyatt 2020) do suggest that (at least some) reclaimed uses of slurs involve the expression of some positive content, other proposals don’t need to. Take for instance Bianchi (2014), according to whom using a slur in a reclaimed way is (i) to evoke the standard derogatory content that slurs typically convey while (ii) expressing at the same time one’s dissociative attitude towards it. In this framework, speakers need not express a positive attitude towards the target class; all they need is to dissociate from the negative attitudes associated with derogatory uses. Put differently, some accounts need not worry about whether the moral permissibility of certain kinds of reclamation is challenged by the WA, because their proposals do not subscribe to the second premise of the WA to begin with (i.e., the idea that at least some reclaimed uses of slurs convey a positive evaluation of the target class).

What about all the other theories according to which reclamation consists in polarity reversal? In the next section, I propose a way to respond to the challenge raised by the WA which is available to possibly any account of reclamation that subscribes to the second premise of the WA.

#### **4. In defense of pride reclamation: a parallel with affirmative action**

In this section I show that, despite appearances, the WA does not necessarily constitute an argument against the moral permissibility of reclamation, if we accept that it can be morally permissible to balance an existing form of injustice by temporarily introducing a countervailing mechanism that—taken out of context—may look like a violation of the norm of equality. To this end, I propose a parallel with a very different phenomenon in a very different domain: the debate on affirmative action. My proposal is to resist the conclusion that it is morally problematic for certain reclaimed uses of slurs to convey unwarranted positive evaluation of the target class, by understanding such uses of epithets as remedies meant to countervail

existing power imbalances. To illustrate this parallel, let us look at affirmative action, typically defined along these lines:

Affirmative action means positive steps taken to increase the representation of women and minorities in areas of employment, education, and culture from which they have been historically excluded. When those steps involve preferential selection—selection on the basis of race, gender, or ethnicity—affirmative action generates intense controversy. (Fullinwider 2013)

Affirmative action gave rise to a lively debate concerning how to interpret it and, above all, how to engage in it, if at all (for a survey on the debate starting from the Seventies, see Fullinwider 2017). A possible way to understand affirmative action is the following: in order to balance an unjust mechanism (negative discrimination), it is morally admissible to introduce a countervailing kind of imbalance (positive discrimination) that is supposed to counteract the initial one over time. Consider the case of an unfair society characterized by systematic gender imbalances where women face undeserved barriers to employment. The employment practice in such a society discriminates on the basis of gender (negative discrimination). One measure that can help to fix this unfair situation is to resort to affirmative action and, in particular, to gender quotas that increase the employment rates of women and try to balance their exclusion (positive discrimination). Such a procedure is meant not only to (partially) balance the past exclusion, but also to start a virtuous circle where a gender-balanced work environment is more likely to avoid the exclusion of women in the future. In a sense (and this point attracted much criticism since the Seventies), affirmative action does not challenge the problem at its roots—i.e., it does not challenge discrimination as a matter of principle, nor does it call into question the fact that gender should not be crucial to whether one gets a job or not. It assumes that the best way to eradicate gender imbalances in the job market—rather than disavowing the very practice of taking the gender of the job candidate into account—is to temporarily modify the valence of discrimination (from negative to positive), in order to achieve anti-discriminatory results in the long run. One could thus say that affirmative action temporarily reproduces the problematic mechanisms it aims to fight, with an opposite valence: if a woman gets hired because of gender quotas (i.e.,



because she's a woman), one could say that the system keeps engaging in employment practices that "discriminate" on the basis of gender. Interestingly, most of the arguments employed to defend affirmative action and quotas are instrumental: they justify the introduction of certain mechanisms that *prima facie* violate the rule of equality in order to fix a previous widespread and systematic injustice. In the words of Goldman: "Thus short-run violations of the rule [of equality] are justified to create a more just distribution of benefits by applying the rule itself in future years" (Goldman 1979: 164-165). Affirmative action is morally justified in as much as it delivers positive results in balancing past discrimination and preventing ongoing and future discrimination. Indeed, a crucial point in the debate around affirmative action concerns the results it produces (for instance, concerning the effects of gender quotas, see Matland 2006; Franceschet and Piscopo 2008; Schwindt-Bayer 2009; Alexander 2012; Kittilson and Schwindt-Bayer 2012; Barnes and Burchard 2013; O'Brien and Rickne 2016).

Let's now go back to the reclamation of slurs. The parallel with affirmative action is meant to suggest that even though certain uses of reclaimed slurs convey the idea that merely belonging to a category makes a person worthy of positive evaluation, such uses are morally permissible in as much as they deliver beneficial results in taking the derogatory content off these terms in the long run. Even if one goes as far as to consider certain kinds of reclamation as short-run violations of the rule of equality (and they may look as such in accounts like Ritchie's), they would be interim solutions to fight the kind of prejudice and discrimination that slurs both express and spread. The question whether reclamation works as a measure to achieve such goals (empowering the victims of discrimination and turning the slur into a less powerful weapon in the long run) is to be investigated on empirical grounds. Experimental studies support the claim that reclamation delivers good results in terms of decreasing the perceived offensiveness of slurring terms and raising the subject's sense of empowerment. According to Galinsky et al. (2003) and Galinsky et al. (2013), the fact that people self-apply slurs has, on the one hand, important effects both on how powerful they feel and how powerful the observers perceive them; on the other hand, the self-application of the slur diminishes the perceived offensiveness of the

term. Moreover, reclamation seems to have eventually succeeded in affecting the derogatory content in some cases: after a process of reclamation, ‘queer’ has today a new non-derogatory use that we can observe in expressions such as ‘Queer Tango’, ‘Queer Film Festival’, ‘Queer Culture’, ‘Queer Studies’, etc. (see Brontsema 2004). What is interesting is that the reclamation of ‘queer’ did not finally turn it into something like a positive slur—i.e., a term conveying a positive value judgement on the target class qua category; rather, it simply made the term non-derogatory. The same might happen for many other reclaimed slurs.

I don’t mean to put too much weight on the empirical claim that reclamation is beneficial; rather, I argue that if it is beneficial as it seems, then accepting the WA does not automatically mean banning reclamation, because—as the case of affirmative action shows—it can be morally permissible to balance an existing form of injustice by temporarily introducing a countervailing mechanism that may *prima facie* seem in violation of the norm of equality, if this measure proves beneficial on empirical grounds. So, if reclamation succeeds in securing beneficial empirical results, then it’s safe from the censorship that could follow from the WA.

## **5. A digression: Black Lives Matter and the myth of reverse racism**

The observations made so far may have further interesting applications beyond the debate on slurs, as this little digression will suggest. On June 6, 2017, the Black commentator and producer Lisa Durden appeared on Fox News’ *Tucker Carlson Tonight* to discuss the issue of why it was legitimate for Black Lives Matter (BLM) to create a ‘Black only’ safe space for the Memorial Day Celebration. Tucker Carlson, a white conservative commentator, accused BLM of being racist: the whole point of BLM is to fight racism and nevertheless they were excluding people on the basis of race. Lisa Durden remarked “you’ve been having ‘white day’ forever, you don’t say the words anymore ‘cause you know it’s politically incorrect, but you’ve had an all-white Oscars, all these movies with all-white actors, movie after movie after movie, (...) and over and over again”. Carlson asked, “I just

have a very simple question for you: if you don't like people excluding others on the basis of their race (...), why are you doing it?". Then again "If you don't like it, why are you perpetuating it?", "Do you think it's racist to exclude people on the basis of their skin color?". Durden answered, "I think it's racist when you've been excluding people for hundreds and hundreds of years and we are forced to come together collectively to celebrate ourselves because you guys won't; you are the largest society: let's be real here". After another exchange, Durden claimed, "Unfortunately, when you have a racist society like America, you force people to come together collectively to make sure that they have a voice". Durden was suspended and then fired from her position of adjunct professor at Essex County College in Newark.

I suggest that the issue discussed in this brief dialogue—was BLM being racist or not?—may be analyzed along the lines illustrated in the previous sections. It may be true that, strictly speaking, excluding people from an event on the basis of their race is something that in an ideal society shouldn't happen, as it would violate the norm of equality; however, in a racist society like the North-American one, it can be morally permissible to temporarily introduce a measure that *prima facie* violates the norm of equality that aims to balance a systemic form of injustice over time.

Similar observations apply more generally to the myth of reverse racism or sexism, according to which members of privileged groups (white people, men, etc.) are victims of the discrimination allegedly perpetuated by underprivileged groups (Black people, women). This kind of attitude suggests that the pursuit of an antiracist and antisexist society may result in the members of the privileged groups ending up being the victim of discrimination. One way to debunk this myth is by underlining how the rich and heterogeneous family of empowering measures (from celebrations in safe spaces to positive uses of slurs) are local and interim solutions to fight power imbalance in the face of systematic and ingrained forms of injustice. There may be a day when Black Lives Matter will not have reasons to exist, nor gender quotas. However, such a day has yet to come and, to say it with a slogan: to make things right, it may be not enough to just do things right.

## 6. Conclusion

In this paper I argued that the moral permissibility of pride reclamation is not automatically ruled out by the WA: as other domains—like political representation and employment practices—show, it can be morally permissible to balance an existing form of negative discrimination by temporarily introducing a countervailing mechanism that seems *prima facie* to violate a norm of equality. This is not to suggest that reclamation and affirmative action are the same thing, but the disanalogies between the two do not seem to threaten my point. An interesting disanalogy worth mentioning is that while reclamation is typically determined and pursued by the group that is discriminated against (and, in some cases, their allies), affirmative action, in contrast, is usually decided from above and it is not mainly pursued or implemented by the discriminated group. Some could thus say that affirmative action involves a bit of paternalism that reclamation lacks. I shall leave the task of defending affirmative action from the charge of paternalism to another occasion, as I do not need the parallel between reclamation and affirmative action to go too far. But it is nevertheless interesting to note that, if one thinks that affirmative action could have non-beneficial effects on the discriminated groups due to its character of a decision from above, this difficulty does not arise for reclamation, that is typically ignited and mainly carried out by the target class and its allies.

To conclude, in this paper I have anticipated and rejected a worry against the moral permissibility of pride reclamation, *i.e.*, reclamatory uses of slurs conveying a positive evaluation of the target. I summarized this criticism in what I dubbed the Warrant Argument. According to the WA, standard and reclaimed uses of slurs make a similar moral error in that both assume an unwarranted connection between descriptive properties (such as being gay, being Italian, being Jewish and so on) and value judgements (being bad/good, being worthy of contempt/pride). The WA could lead us to condemn or ban pride reclamation, just like we do with derogatory uses of slurs, but I have proposed a parallel with the mechanisms of affirmative action that provides a way to resist this conclusion. The case of affirmative action shows that it can be morally permissible to balance an existing form of injustice by temporarily introducing a countervailing mechanism that

prima facie seems to violate the norm of equality, if the measure proves effective in fighting oppression. The analogy with affirmative action suggests that, even if the WA were right, it wouldn't constitute an argument against the moral permissibility of pride reclamation. This strategy to defend the moral permissibility of reclamation from the WA is not the most direct one, as one could simply reject one of WA's premises. The advantage of this longer journey is that by granting more to the WA, its conclusion should be appealing to a larger audience.

### Acknowledgements

Many thanks to Dan Zeman for putting together such a rich special issue and a wonderful workshop, Value in language (March 29-31, 2021, Slovak Academy of Sciences). Thanks to all the participants for their insights. In particular, I shall thank for their suggestions Victor Carranza, Nils Franzén, Leopold Hess, Robin Jeshion, Zuzanna Jusińsk, Stefano Predelli, Andrés Soria Ruiz, Pekka Väyrynen, Julia Zakkou and Dan Zeman. I have also benefitted from the insights of Claudia Bianchi, Laura Caponetto, Teresa Marques and Dan Zeman, commenting on earlier versions of this work.

### References

- Alexander, Amy C. 2012. "Change in Women's Descriptive Representation and the Belief in Women's Ability to Govern: A Virtuous Cycle." *Politics & Gender* 8 (4): 437–64. <https://doi.org/10.1017/S1743923X12000487>
- Allan, Keith. 2015. "When Is a Slur not a Slur? The Use of *nigger* in 'Pulp Fiction'". *Language Sciences* 52: 187–99. <https://doi.org/10.1016/j.langsci.2015.03.001>
- Anderson, Luvell. 2018. "Calling, Addressing, and Appropriation." In *Bad Words: Philosophical Perspectives on Slurs*, edited by David Sosa, 15–37. Oxford, Oxford University Press. 10.1093/oso/9780198758655.003.0002
- Anderson, Luvell, and Ernie Lepore. 2013. "Slurring Words." *Noûs* 47 (1): 25–48. <https://doi.org/10.1111/j.1468-0068.2010.00820.x>
- Asim, Jabari. 2007. *The N Word: Who Can Say It, Who Shouldn't, and Why*. New York-Boston: Houghton Mifflin Company.
- Bailey, Guy, John Baugh, Salikoko S. Mufwene, and John R. Rickford (eds.). 1998. *African-American English: Structure, History and Use* (1 edition). London, New York: Routledge.

- Barnes, Tiffany D., and Stephanie M. Burchard. 2013. "'Engendering' Politics, The Impact of Descriptive Representation on Women's Political Engagement in Sub-Saharan Africa." *Comparative Political Studies* 46 (7): 767–90. <https://doi.org/10.1177/0010414012463884>
- Bianchi, Claudia. 2014. "Slurs and Appropriation: An Echoic Account." *Journal of Pragmatics* 66: 35–44. <https://doi.org/10.1016/j.pragma.2014.02.009>
- Brontsema, Robin. 2004. "A Queer Revolution: Reconceptualizing the Debate over Linguistic Reclamation." *Colorado Research in Linguistics* 17: 1–17. <https://doi.org/10.25810/dky3-zq57>
- Burnett, Heather. 2020. "A Persona-Based Semantics for Slurs." *Grazer Philosophische Studien* 97 (1): 39–62. <https://doi.org/10.1163/18756735-09701004>
- Camp, Elisabeth. 2013. "Slurring Perspectives." *Analytic Philosophy* 54 (3): 330–49. <https://doi.org/10.1111/phib.12022>
- Camp, Elisabeth. 2018. "A Dual Act Analysis of Slurs." In *Bad Words: Philosophical Perspectives on Slurs*, edited by David Sosa, 29–59. Oxford, Oxford University Press. [10.1093/oso/9780198758655.001.0001](https://doi.org/10.1093/oso/9780198758655.001.0001)
- Childs, Sarah, and Mona Lena Krook. 2009. "Analysing Women's Substantive Representation: From Critical Mass to Critical Actors." *Government and Opposition* 44 (2): 125–45. <https://doi.org/10.1111/j.1477-7053.2009.01279.x>
- Croom, Adam. 2011. "Slurs". *Language Sciences* 33 (3): 343–58. <https://doi.org/10.1016/j.langsci.2010.11.005>
- Croom, Adam. 2013. "How to Do Things with Slurs: Studies in the Way of Derogatory Words". *Language & Communication* 33 (s): 177–204. <https://doi.org/10.1016/j.langcom.2013.03.008>
- Croom, Adam. 2014. "Spanish Slurs and Stereotypes for Mexican-Americans in the USA: A Context-Sensitive Account of Derogation and Appropriation". *Sociocultural Pragmatics* 2 (2): 145–79. <https://doi.org/10.1515/soprag-2014-0007>
- Dahlerup, Drude, and Lenita Freidenvall. 2010. "Judging Gender Quotas: Predictions and Results." *Policy & Politics* 38 (3): 407–25. [10.1332/030557310X521080](https://doi.org/10.1332/030557310X521080)
- Darwall, Stephen L. 1977. "Two Kinds of Respect." *Ethics* 88 (1): 36–49. <https://doi.org/10.1086/292054>
- Dworkin, Gerald. 2017. "Paternalism." In *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), edited by Edward N. Zalta. URL = <https://plato.stanford.edu/archives/spr2017/entries/paternalism/>
- Franceschet, Susan, and Jennifer M. Piscopo. 2008. "Gender Quotas and Women's Substantive Representation: Lessons from Argentina." *Politics & Gender* 4 (3): 393–425. <https://doi.org/10.1017/S1743923X08000342>

- Fullinwider, Robert. 2017. "Affirmative Action." In *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), edited by Edward N. Zalta. URL = <https://plato.stanford.edu/archives/sum2017/entries/affirmative-action/>
- Galinsky, Adam D., Kurt Hugenberg, Carla Groom, and Galen V. Bodenhausen. 2003. "The Reappropriation of Stigmatizing Labels: Implications for Social Identity." In *Identity Issues in Groups*, edited by Jeffrey T. Polzer, 221–56. Emerald Group Publishing Limited. [https://doi.org/10.1016/S1534-0856\(02\)05009-0](https://doi.org/10.1016/S1534-0856(02)05009-0)
- Galinsky, Adam D., Cynthia S. Wang, Jennifer A. Whitson, Eric M. Anicich, Kurt Hugenberg, and Galen V. Bodenhausen. 2013. "The Reappropriation of Stigmatizing Labels: The Reciprocal Relationship between Power and Self-labeling." *Psychological Science* 24 (10): 2020–9. <https://doi.org/10.1177/0956797613482943>
- Herbert, Cassie. 2015. "Precarious Projects: the Performative Structure of Reclamation." *Language Sciences* 52: 131–138. <https://doi.org/10.1016/j.langsci.2015.05.002>
- Hess, Leopold. forthcoming. "Slurs: Semantics and Pragmatics Theories of Meanings." In *The Cambridge Handbook of Philosophy of Language*, edited by Piotr Stalmaszczyk. Cambridge University Press.
- Jeshion, Robin. 2013a. "Expressivism and the Offensiveness of Slurs." *Philosophical Perspectives* 27 (1): 231–59. <https://doi.org/10.1111/phpe.12027>
- Jeshion, Robin. 2013b. "Slurs and Stereotypes." *Analytic Philosophy* 54 (3): 314–29. <https://doi.org/10.1111/phib.12021>
- Jeshion, Robin. 2020. "Pride and Prejudiced: On the Appropriation of Slurs." *Grazer Philosophische Studien* 97 (1): 106–37. <https://doi.org/10.1163/18756735-09701007>
- Kennedy, Randall. 2002. *Nigger: The Strange Career of a Troublesome Word*. New York: Vintage.
- Kittilson, Miki Caul, and Leslie A. Schwindt-Bayer. 2012. *The Gendered Effects of Electoral Institutions: Political Engagement and Participation*. New York: Oxford University Press. [10.1093/acprof:oso/9780199608607.001.0001](https://doi.org/10.1093/acprof:oso/9780199608607.001.0001)
- Kleinman, Sherryl, Matthew B. Ezzell, and Corey Frost. 2009. "Reclaiming Critical Analysis: The Social Harms of 'Bitch'." *Sociological Analysis* 3 (1): 46–68.
- Marques, Teresa. 2017. "Pejorative Discourse Is Not Fictional." *Thought: A Journal of Philosophy* 6 (4): 250–60. <https://doi.org/10.1002/tht3.258>
- Marques, Teresa, and Manuel García-Carpintero. 2020. "Really Expressive Presuppositions and How to Block Them." *Grazer Philosophische Studien* 97 (1): 138–58. <https://doi.org/10.1163/18756735-09701008>

- Matland, Richard E. 2006. "Electoral Quotas: Frequency and Effectiveness." In *Women, Quotas and Politics*, edited by Drude Dahlerup, 275-92. London, Routledge. [10.1080/00344890701363227](https://doi.org/10.1080/00344890701363227)
- Miščević, Nenad, and Julija Perhat. 2016. *A Word Which Bears a Sword: Inquiries into Pejoratives*. Zagreb, Kruzak.
- Murray, Rainbow. 2010. "Second among Unequals? A Study of Whether France's 'Quota Women' Are Up to the Job." *Politics & Gender* 6 (1): 93-118. <https://doi.org/10.1017/S1743923X09990523>
- Nunberg, Geoffrey. 2018. "The Social Life of Slurs." In *New Work on Speech Acts*, edited by Daniel Fogal, Daniel W. Harris, and Matt Moss, 237-95. Oxford: Oxford University Press. [10.1093/oso/9780198738831.001.0001](https://doi.org/10.1093/oso/9780198738831.001.0001)
- O'Brien, Diana Z., and Johanna Rickne. 2016. "Gender Quotas and Women's Political Leadership." *American Political Science Review* 110 (1): 112-26. <https://doi.org/10.1017/S0003055415000611>
- Popa-Wyatt, Mihaela. 2020. "Reclamation: Taking Back Control of Words". *Grazer Philosophische Studien* 97 (1): 159-76. <https://doi.org/10.1163/18756735-09701009>
- Richard, Mark. 2008. *When Truth Gives Out*. Oxford: Oxford University Press.
- Ritchie, Katherine. 2017. "Social Identity, Indexicality, and the Appropriation of Slurs." *Croatian Journal of Philosophy* 17 (2): 155-80. [hrcak.srce.hr/195051](https://doi.org/10.1563/cjphil/17.2.155)
- Schwindt-Bayer, Leslie A. 2009. "Making Quotas Work: The Effect of Gender Quota Laws on the Election of Women." *Legislative Studies Quarterly* 34 (1): 5-28. <https://www.jstor.org/stable/20680225>
- Stanley, Jason. 2015. *How Propaganda Works*. Princeton: Princeton University Press.
- Tirrell, Lynne. 1999. "Derogatory Terms: Racism, Sexism, and the Inferential Role Theory of Meaning". In *Language and Liberation: Feminism, Philosophy, and Language*, edited by Christina Hendricks, and Kelly Oliver, 41-77. Albany (NY): SUNY Press.
- Thompson, Nicole Akoukou. 2013. "John Leguizamo & Kanye West use Re-appropriation to Change Perceptions." *Latin Post*, 11 November. <http://www.latin-post.com/articles/3547/20131111/>
- Zeman, Dan. 2021. "A Rich-Lexicon Theory of Slurs and Their Uses." *Inquiry*, <https://doi.org/10.1080/0020174X.2021.1903552>



# Slur Reclamation – Polysemy, Echo, or Both?

Zuzanna Jusińska\*

Received: 18 November 2020 / 1<sup>st</sup> Revised: 14 May 2021 /  
2<sup>nd</sup> Revised: 9 July 2021 / Accepted: 8 August 2021

*Abstract:* This paper concerns the topic of slur reclamation. I start with presenting two seemingly opposing accounts of slur reclamation, Jeshion’s (2020) Polysemy view and Bianchi’s (2014) Echoic view. Then, using the data provided by linguists, I discuss the histories of the reclamation of the slur ‘queer’ and of the n-word, which bring me to presenting a view of reclamation that combines the Polysemy view and Echoic view. The Combined view of slur reclamation proposed in this paper postulates meaning change while fleshing out the pragmatic mechanisms necessary for it to occur.

*Keywords:* Meaning change; pragmatics; reclamation; semantics; slurs.


## 1. Introduction

The goal of this paper is to explore the topic of slur reclamation. It seems that explaining slur reclamation *via* pragmatic mechanisms competes with accounts which posit a semantic ambiguity between the derogatory and the reclaimed slur. I argue that these views are not rivals; they complement each other. I claim that it is impossible to explain meaning change

---

\* University of Warsaw

 <https://orcid.org/0000-0002-9490-6799>

 Krakowskie Przedmieście 26/28, 00-927 Warszawa, Poland.

 [z.jusinska@uw.edu.pl](mailto:z.jusinska@uw.edu.pl)



without appealing to pragmatic mechanisms, especially in the case of slur reclamation, given the socio-political motivation of this process.

In sections 2 and 3 I present two accounts of slur reclamation, Jeshion's (2020) Polysemy view and Bianchi's (2014) Echoic view. Section 4 consists of a discussion of the histories of reclaiming the slur 'queer' and of the n-word. In the 5<sup>th</sup> section, I propose a view of slur reclamation that combines the Polysemy and Echoic accounts.

## 2. Jeshion's Polysemy view

Jeshion (2020) distinguishes two most common variants of slur reclamation—pride reclamation (such as in the reclamation of 'queer') and insular reclamation (such as in the reclamation of the n-word). She defines pride reclamation as “the reclamation of a pejorative representation through processes in which the representation is accompanied by expressions of pride for being in the group or the targeted object, and the representation is presented publicly as an apt way to reference the group” (2020, 107); and insular reclamation as “the reclamation of a pejorative representation through processes in which use of the representation dominantly functions to express and elicit camaraderie among target members in the face of and to insulate from oppression, and the representation is not presented publicly as an apt way for out-group members to reference target group members” (2020, 107). In short, to use these paradigmatic examples<sup>1</sup>, reclaiming 'queer' expresses pride in being not-cisgender/not-heterosexual and presents this word to cisgender, heterosexual people as an appropriate way of referring to not-cisgender/not-heterosexual people, while reclaiming the n-word expresses camaraderie and solidarity between Black people in defiance of racism and this word is not presented to non-Black people as an appropriate way of referring to Black people.

Jeshion (2020) notes that reclamation is a complex linguistic and social process which involves numerous individual and collective acts performed and interpreted within the relevant communities and that this process

---

<sup>1</sup> Among other examples used by Jeshion are 'Black' for pride reclamation and 'bitch' for insular reclamation.

extends through a long period of time. She characterizes the diachronic structure of the process of slur reclamation, as it is often characterized for other instances of linguistic change, as having four stages:

- I) Preliminary state: the word is governed by linguistic conventions C regarding its meaning, pragmatic use, primary associations.
- II) Acts of linguistic creativity and innovation: speakers use the word in novel ways, departing from C, sometimes with the deliberate aim to effect change, sometimes not.
- III) Acts of imitation and diffusion: speakers imitate the novel uses or key aspects of them.
- IV) End result: the word has come to be governed by new linguistic conventions C'≠C; the word may still retain its former conventions C, becoming polysemous, or C may be supplanted by C'. (Jeshion 2020, 108)

Jeshion characterizes *initial reclamatory acts* as those which ignite the reclamation process (stage 2) and *secondary reclamatory acts* as imitative and parasitic on the previous ones (stage 3). During the reclamation process slurs are polysemous, often for a long time, and they retain the linguistic conventions encoding derogation while simultaneously acquiring the non-derogatory ones. Initial reclamatory acts consist of speakers intentionally breaking and altering the linguistic conventions in order to change the oppressive social norms justifying and manifested by the slur. The speakers imitating the initial reclaimers do not necessarily have such intentions but their uses are still a part of the reclamation process and of the emancipatory movement. Jeshion states that “aiming to break linguistic conventions to shift oppressive social norms that are manifested and perpetuated by linguistic representations is a key ingredient to acts of reclamation” (2020, 111), taking the speakers’ intentions to play a big part in the initiation of the reclamation process. Usually, in the initial acts of reclamation the speakers aim to undermine the slur’s conventional function as a weapon by hijacking it and using it in a positive way.

Jeshion claims that initial pride reclamatory acts and initial insular reclamatory acts differ. She writes that in the pride ones the members of the target group intentionally and consciously use the slur in a novel way to

change the dominant negative attitudes towards them. Using the slur the “speakers self- and group-reference while overtly manifesting an attitude of pride for being in the target group” (2020, 121). These acts are acts of self- and group-affirmation—the slur becomes the group identity-label. On the other hand, in insular reclamatory acts the target group members use a slur in a novel way but not necessarily with a conscious goal of responding to or transforming the dominant negative attitudes present in the society. Jeshion claims that even though the initial acts are similar to ordinary apolitical in-group uses of unreclaimed slurs (similar to mock insults between friends), “later acts quickly become intentionally political” (2020, 121). According to Jeshion, acts of imitation are often performed with an awareness of their political power and aimed at achieving a sense of solidarity. In opposition, pride reclamatory acts are direct and sincere and not mocking or ironic. Jeshion claims that while such acts might combine manifesting pride with expressing disdain or mockery of bigotry, “the latter [is] not necessary, and often non-existent” (2020, 121, footnote 21). She writes that it is the first-order use of the slur with the positive polarity that secures the linguistic change. Meanwhile, insular reclamatory acts mock the derogation present in the slur uses and ridicule it into a term manifesting camaraderie.

According to Jeshion, the semantic change achieved through the process of slur reclamation begins with acts of linguistic innovations generating meaning-transformations. She claims that initial pride reclamatory acts achieve amelioration by connecting the slur with paralinguistic cues and positive associations: “speakers express pride or group-self-respect through overt statements, but also intonation, gesture, body language, visibility when the norm is the closet or silence” (2020, 125). To put it differently, in pride reclamatory acts the speakers introduce transformed slur meanings (to be secured by the widespread use and conventionalisation) by using the slur while communicating (directly or indirectly) pride in belonging to the target group. Meanwhile, initial insular reclamatory acts involve verbal irony, which can be emphasized by amelioration via paralinguistic cues and positive associations. In short, initially insular reclamatory acts are ironic uses of the slur aimed at communicating camaraderie. After such uses become widespread and conventionalized the slur “shifts polarity and becomes a social deictic for communicating camaraderie” (2020, 125).

The difference between pride and insular reclamation is crucial to Jeshion's description of secondary uses of reclaimed slurs. She claims that while this may differ between particular slurs, in pride reclamation secondary out-group use becomes permissible and in insular reclamation it is generally prohibited. In the case of pride reclamation initial acts present the slur as a group-adopted identity label which, along the normalization through in-group imitation, amounts to the target group tacitly authorizing out-groups to use the reclaimed slur as an appropriate way to refer to the target-group. In the case of insular reclamation, the target group does not adopt the slur as a group identity-label and the widespread in-group uses do not authorize out-group use. On the contrary, according to Jeshion, in-group uses function to communicate camaraderie in the shared experience of discrimination towards the target group and therefore prohibit out-group uses.

### 3. Bianchi's Echoic view

Bianchi (2014) offers an echoic account of slur reclamation. According to her view, when the members of the target group use the relevant slur in a reclamatory way, they echo the derogatory uses and manifest their dissociation from the derogatory contents. This view is supposed to account for the fact that appropriative uses of a slur are typically available only for the members of the targeted group, although they can be extended to selected non-members in highly regulated situations.

What is worth noting is that Bianchi distinguishes between two types of contexts of in-group non-derogatory uses of slurs that are supposed to demarcate the group and show a sense of intimacy and solidarity. These are *the friendship contexts*, where there is no conscious political intent, and *the appropriation contexts*, where target groups reclaim the use of the slur as a deliberate socio-political action or artists belonging to the group attempt appropriation as a way of subverting the oppressive socio-cultural norms. Bianchi's distinction differs from Jeshion's (2020) distinction between pride reclamation and insular reclamation—although, on the face of it, the latter resembles Bianchi's friendship contexts. While both Bianchi's and Jeshion's distinctions focus on the speaker's intention, for Bianchi the friendship and appropriative contexts differ in terms of the lack or presence of conscious

socio-political intent. Most importantly, Jeshion distinguishes between the reclamation of slurs claiming that, e.g., ‘queer’ falls into the category of pride reclamation while the n-word falls into the category of insular reclamation, whereas Bianchi distinguishes between friendship and appropriative *uses* of slurs. Throughout the paper Bianchi rarely differentiates between these two contexts and refers to reclamatory uses as *community uses* of slurs.

What is crucial for Bianchi’s view is that it accounts for non-derogatory uses of slurs (including reclamation) without postulating meaning change. The echoic account of slur reclamation is based on the echoic uses of language as they are defined in the Relevance Theory introduced by Wilson and Sperber (1986). According to the Relevance Theory, we can distinguish between descriptive and interpretive uses of language. A descriptive use of an utterance or a thought represents a state of affairs in the world, while an interpretive use represents the (actual or possible) utterance/thought of another person concerning a state of affairs. An example of an interpretive use of language is an indirect speech report. Echoic uses are a subset of interpretive uses in which a speaker both represents an attributed utterance/thought and informs (e.g., *via* intonation, facial expressions or other context cues) the hearer of their attitude towards that utterance or thought. Ironic uses are those echoic uses in which the speaker’s attitude towards the attributed content is dissociative. Bianchi claims that in the case of ironic uses the speaker expresses a dissociative attitude either towards an actual or possible utterance/thought attributed to another person or towards (cultural, moral, social, etc.) expectations and norms.

Bianchi proposes an echoic account of slur reclamation—reclamatory uses echo derogatory uses in ways that manifest their dissociation from the offensive contents expressed or conveyed by slurs. She claims that often these are ironic uses in which the speaker attributes utterances or thoughts to others in order to express a critical attitude. Bianchi emphasizes that this attitude might differ between speakers, ranging from “playful puzzlement to powerful condemnation, from joyful mockery to harsh rejection, and so on” (2014, 40). Cepollaro (2020) further develops Bianchi’s echoic view and offers insightful remarks on the power of irony writing that “[i]n ridiculing and mocking the bigot’s perspective by using their own words, the

speaker puts herself in a position of superiority: she steals a weapon and refuses to surrender to discrimination and prejudice; she refuses to be just a suffering victim or a powerless witness of hate speech and instead resists by subverting linguistic conventions” (2020, 90–91).

Bianchi provides an example of friendship context—in which members of the target-group use the slur non-offensively in order to express a sense of closeness and solidarity with no conscious socio-political intent—where two gay friends, Al and Bob, talk about a new colleague, Tom, and Al utters (1):

(1) I’m sure Tom is a faggot.

In this scenario, Al uses the slur ‘faggot’ to echo a representation with a conceptual content—“a cultural, moral or social norm stating that homosexuals deserve derision or contempt” (Bianchi 2014, 40). Al communicates his own dissociative attitude towards this homophobic norm and suggests that the idea that gay people deserve contempt is false, stupid, inappropriate, bad, shameful, etc.

What may seem like a problem for Bianchi’s account is that in such non-derogatory uses as in (1), the speaker does assert something. By uttering (1), Al not only mocks the homophobia represented in the slur ‘faggot’, but also represents a state of affairs such that he is sure that Tom is gay. Bianchi claims that indeed in such uses the speaker commits oneself to the assertion of the sentence with a neutral counterpart in the place of the slur, but not to the offensive content expressed or conveyed by the slur ‘faggot’. Furthermore, we could assume that Al is not echoing a concept but only a constituent of concept—its derogatory component.

One of the last issues that an account of slur reclamation needs to cover is the difference between in-group and out-group uses. Bianchi explains the fact that reclamatory uses are usually only available to the members of the target-group in the following way: “an ironical use requires a context in which the dissociation from the echoed offensive content is clearly identifiable: *ceteris paribus*, in-group membership is *per se* strong evidence that the exchange takes place in such a context” (2014, 42). While out-groups can have dissociative attitudes towards the slur’s derogatory context, it is impossible for them to undoubtedly make this attitude manifest. Even when

their interlocutors are aware of their attitudes and opinions, anyone who overhears the utterance could take it to be a derogatory one. On the other hand, Bianchi claims that appropriated uses of slurs may extend to out-groups. She claims that selected speakers and highly controlled conditions can create contexts in which the out-groups' dissociation from the derogatory contents is clear. Bianchi claims that this was the case for the word 'queer'—the LGBT+ community authorized the academic community to use this term in an appropriated way.

It needs to be noted that Bianchi does realize that uses of some reclaimed slurs are no longer echoic. She claims that for words such as 'gay' or 'queer' the reclamation process is over and, when that happens, we can say that the meanings of these words has changed (or that the words no longer convey offense). While her account focuses on the linguistic mechanisms of the particular reclamatory uses of slurs and not on reclamation as a process, she does note that when the practice of reclamatory uses "is sufficiently widespread it may extend also to selected out-groups, and affect—diachronically—the slur meaning (expressed or conventionally conveyed)" (2014, 43).

#### 4. The reclamation of 'queer' and of the n-word

In this section I will discuss some data concerning the reclamation of the slur 'queer' and of the n-word provided by Brontsema (2004) and Rahman (2012). Brontsema (2004) provides a linguistic account of reclamation focusing on the specific case of the term 'queer'. She cites Chen's definition: "The term 'reclaiming' refers to an array of theoretical and conventional interpretations of both linguistic and non-linguistic collective acts in which a derogatory sign or signifier is consciously employed by the 'original' target of the derogation, often in a positive or oppositional sense" (1998, 130). I will briefly discuss the history of the reclamation of the word 'queer' provided by Brontsema.

During the 1980s and the early 1990s, the LGBT+ community started to reclaim the term 'queer' which was then the most popular and harmful slur for gay and trans people. The homophobia in the AIDS activism and the increase in anti-gay crimes lead to launching of several activist groups



including Queer Nation, some of whose members were responsible for the famous “Queers Read This” flyers handed out at the 1990 Gay Pride Parade in New York City (Rand 2014). The flyer urged its readers to take a stand against homophobic and heterosexist institutions, to reclaim the word ‘queer’ as a form of resistance and join forces under its banner. Reclaiming the word ‘queer’ was in itself a radical act of highlighting homophobia in order to fight it. The “Queers Read This” flyer underlined the need for direct action and objected to the assimilationist strategies with straightforward statements such as ‘Straight people are your enemy.’ Reclaiming the term ‘queer’ set out to unite people of non-normative sexualities and genders, and it was not meant to be used as a synonym for gay and lesbian (Brontsema 2004).

Brontsema claims that there were several uses of the term ‘queer’ that coexisted (at the time of writing) and that it is not the case that there are only positive in-group uses and negative out-group uses. She discusses the use of the term ‘queer’ by self-identified queers, in which the term is used inclusively and in opposition to the essentializing ‘gay’ and ‘lesbian’, and in which it can be understood as more of an anti-identity than identity. Such use is similar to how the reclaimed meaning of ‘queer’ was intended in the early 1990s, but the larger society generally failed to understand the nuances of the term and uses it as a synonym of gay and lesbian. Another use of ‘queer’ is the one appearing in popular television series such as “Queer as Folk” and “Queer Eye for the Straight Guy”, focusing on gay men, where it is used rather as a trendy synonym for ‘gay’. The reclamation process did not eliminate the derogatory use of ‘queer’ and during the reclamation process the word continued to be used pejoratively. The last use of ‘queer’, which also diverges from the radical meaning intended in the initial acts of reclamation, is the contemporary most common meaning of the word—an umbrella term for lesbian, gay, bisexual and transgender people. Today, with the rising awareness of many different sexualities and gender identities, the acronym extends to LGBTQIAP+<sup>2</sup> or, in a shorter version, to LGBTQ+ where the ‘queers’ are included in the acronym rather than

---

<sup>2</sup> Meaning: lesbian, gay, bisexual, transgender, queer, intersex, asexual, pansexual. The ‘+’ sign is included to emphasize the fact that there are more ways to identify beyond the acronym.

equated with it. Nevertheless, the term ‘queer’ is now most commonly understood as “of, relating to, or being a person whose sexual orientation is not heterosexual and/or whose gender identity is not cisgender” (definition taken from the online entry for ‘Queer’ in the Merriam Webster dictionary at <https://www.merriam-webster.com/dictionary/queer>).

The history of the term ‘Black’ is similar to that of the term ‘queer’. Brontsema writes that “[t]he history of ‘black’ shows that revolutionary intent does not predetermine the future of a word, that intent can be betrayed even when a word is said to be ‘reclaimed’.” (2004, 11). ‘Black’ was intended to be confrontational, revolutionary and reevaluating Blackness, but when white people became familiar with this term it became just a substitute for the no longer accepted by the Black community term ‘Negro’. As in the history of the term ‘queer’, “[t]he original energy of black was betrayed and subsequently died as it was not used with the same vital radicalism. Instead of forcing racists to confront their hatred and speak it out loud, their racism was simply given a new mask to wear.” (2004, 11).

Rahman (2012) discusses the history of the n-word. While the early use of the word—dating back to 16th century, borrowed from the Spanish and Portuguese slave traders who used the word ‘negro’ meaning ‘black’—was a relatively neutral referential term for Black people used by white people, it became a racist slur during the 19th century. The transformation of the n-word into a racial slur came at the time of the movement for the abolition of slavery and the increase of numbers of free African Americans. Rahman claims that the in-group uses of the n-word developed within the slave community. This variation of the n-word in the African American community can be distinguished by its pronunciation—in African American English the form of the word ends in a schwa, without /r/. Rahman notes that the “social meanings developed among the Africans (...) reflected a view in which they saw themselves as survivors and as humans whose freedom and dignity had been assaulted” (2012, 146). During the time of slavery, the n-word had developed social meanings related to survival—it was a term that Africans used to refer to themselves and others in the struggle to survive and using it emphasized the identity of the speaker as participating in the culture of survival.

In other words, the reclaimed n-word has a core meaning which has been established through generations which relates to survival. Rahman notes that additional but related attitudinal aspects of meaning can layer over the core, such as the solidarity meaning which “emerges through common understanding and shared experiences related to survival” (2012, 155), or the hip-hop community use of the n-word which “underlies projection of an identity that directly and overtly rejects racist uses of [the n-word] while declaring self-pride and independence” (2012, 159). It is worth noting that while the positive uses of the n-word were present in the African American community long before the emergence of the hip-hop community uses, in the last decades of the 20th century these uses became much more widespread because young African Americans in the hip-hop community took ownership of the racist n-word and transformed it into their own positive version of the n-word ending in ‘-a’ instead of ‘-er’.

The reclamation processes are complex and nuanced, both in the case of reclaimed slurs used only by in-groups and in the case of reclaimed slurs open to out-groups. In the next section I will present a view of slur reclamation that accounts for the data provided by Brontsema (2004) and Rahman (2012).

## 5. Combining the Echo and Polysemy views

I want to propose a view of slur reclamation that accounts for the meaning-change without omitting the crucial pragmatic steps, which is motivated by the histories of reclaimed slurs presented in section 4. I take Jeshion’s (2020) and Bianchi’s (2014) accounts of reclamation to be insightful descriptions of different stages of the reclamation process; however, neither account tells the whole story of slur reclamation. Treating the Polysemy and Echoic views as rivals is mistaken and combining the two can account for the linguistic evidence concerning slur reclamation.<sup>3</sup> Furthermore, I believe that such an account can be useful for studying semantic change in general.

---

<sup>3</sup> I am not alone in this stance (see Cepollaro 2020 for a view of slur reclamation that incorporates both the mechanism of echo and polysemy). I agree with Cepollaro

The main reason why Bianchi's (2014) view is not enough to explain slur reclamation is the fact that it does not account for meaning change, although she does note that a widespread echoic use of a slur can give rise to polysemy. Jeshion argues against the echoic view writing that "because ironic echoic utterances of slurs leave intact slurs' weapon meanings, they do not enact any linguistic innovation, and consequently the theory doesn't explain the mechanisms by which pride- and insular-reclaimed slurs become neutralized" (2020, 134). To me, the fact that Bianchi's account cannot account for meaning change is not an objection against her view, but rather a motivation for incorporating her analysis of reclamatory uses into a bigger picture of slur reclamation. Bianchi herself notices that some reclaimed slurs no longer have an ironic component, but she does not explain how that happens. There seem to be two basic possibilities compatible with the echoic account: either the irony gets conventionalized as part of the slur's meaning; or the non-derogatory use of the slur becomes so widespread that the need for distorting the derogation vanishes. Both seem unsuited for explaining the subversive reclamatory acts of self- or group-identification with the slur. However, the echoing of the slur's derogatory content is a necessary step in slur reclamation.

A disadvantage of Jeshion's view, which can be generalized to polysemy views as such, concerns neglect of the pragmatic mechanisms necessary for the new linguistic conventions to emerge. This neglect amounts to Jeshion's (2020) inadequate description of the reclamation of 'queer' and 'Black'. As it was shown in the previous section, the actual initial intent was to use these words in a radical confrontational manner which was to be achieved through keeping the derogatory content of the slur detectable while subverting it and thereby disarming it. To consider this as simply an act of linguistic innovation omits certain crucial features of the initial reclamatory acts. Again, this is not an argument against Jeshion's view of reclamation, but a reason to refine and develop it. Acknowledging the meaning-transformational power of pragmatic mechanisms can also help with accounting for the fact that the outcome of the slur reclamation can often differ from the intent behind the initial reclamatory acts—as in the cases of 'queer' and

---

that while the reclamation process starts with echo, the echoic framework cannot explain how the global meaning of a reclaimed slur changes.

‘Black’. In the following paragraphs, I present an account of slur reclamation which starts from Jeshion’s view and enriches it with pragmatic mechanisms.

Following Jeshion, I want to characterize the stages of slur reclamation explaining at the same time how the slurs come to have new meaning. These stages are characterized as follows:

- (1) Preliminary state: the slur is governed by linguistic conventions *C* regarding its meaning, pragmatic use, primary associations.
- (2) Echoic uses of the slur: in-group speakers echo the slur’s derogatory content manifesting their dissociative attitudes towards it.
- (3) Self- or group-identification: in-group speakers self- or group-identify with the echoic use of the slur. Initiation of a new linguistic convention *C*’.
- (4) Acts of imitation and diffusion: in-group speakers imitate the self- or group-identification uses or key aspects of them. Securing of the linguistic convention *C*’.
- (5) Out-group recognition: the linguistic convention *C*’ reaches the out-group. Various possibilities: (i) *C*’ is adopted by out-groups; (ii) *C*’ is recognized but not adopted by out-groups; (iii) *C*’ is transformed into another linguistic convention *C*’’ by the out-groups.
- (6) Possible end results: (i) polysemy—different linguistic conventions coexist; (ii) replacement—*C*’ or *C*’’ supplants *C*.

Before explaining this process, let me note that at each stage the necessary action for completing the reclamation process may not happen, due to various reasons such as the existence of power imbalances in society, the invisibility of the target group, legislation discriminating against the target group, etc. Moreover, reclamation is a complex process which requires various contextual as well as cultural, social, and political conditions to be successful.<sup>4</sup>

In the first stage the slur has derogatory content and is used to harm the target group. That is the starting point of any reclamation process. In the second stage I make use of Bianchi’s Echoic approach—the members of the target group start using the slur in an echoic way. The speakers, often angry because of being called with the slur and/or disagreeing with the

---

<sup>4</sup> See Herbert (2015) for an insightful analysis of how risky the attempts at reclamation are and what negative consequences they might bring about.

discrimination it connotes, echo the derogatory content of the slur expressing a dissociative attitude towards it. What is important to note is that the dissociative attitudes can vary across speakers and range from mere ridiculing to hateful contempt. During the second stage the meaning of the slur remains intact, as the dissociation from the derogation is achieved by pragmatic mechanisms.

The third step is the self- or group-identification with the echoic use of a slur—the target group members already familiar with the echoic uses of the slur take these uses (mocking, condemning, denouncing the derogatory content of the slur) and associate themselves with them. This step is what makes the reclamatory acts revolutionary, and this is indeed what happened in the cases of ‘queer’ and the n-word. The acts initiating the new meaning of ‘queer’ were acts of displaying rage towards homophobia and “queerbashing”. The acts initiating the new meaning of the n-word in the 1800s were acts of displaying solidarity in the face of oppression. The self- or group-identification with echoing the slur’s derogatory content introduces a new linguistic convention which includes the slur’s new subversive meaning. I take this positive act to be what makes reclamation a case of meaning change, as it is no longer only a dissociation from the derogatory content of the word but rather an introduction of a new content associated with the word.<sup>5</sup> While the mechanism of self- or group-identification is similar to what Jeshion (2020) describes as the Identity Ownership<sup>6</sup> feature of reclamation, the difference lays in what the speakers take to be a part of their identity. On my account, that is the echoing of the derogatory content of a slur, and not simply the reversed-polarity version of the slur.<sup>7</sup>

<sup>5</sup> I follow Jeshion in taking the “novel first-order uses of the slur” (2020, 134) to be necessary for the introduction of a new linguistic convention or, in other terms, a new local meaning.

<sup>6</sup> Jeshion (2020) takes Identity Ownership to be one of the central features of reclamation. She claims that the speakers use the reclaimed slur as an identity-label “as a means to socially group self-define on their own terms” (2020, 122) and at the same time they use the reclaimed slur “as a means to reverse derogating social attitudes and norms on the group” (2020, 123).

<sup>7</sup> Polarity Reversal is another central feature of reclamation on Jeshion’s (2020) account: “speakers use representations that standardly have a negative polarity to communicate a positive polarity” (2020, 122).

The fourth step secures the linguistic convention introduced in the third step. If the new use of the slur becomes widespread enough among the target-group members, the new local meaning is secured. It is important to note here that there might be more than one new meaning introduced in the third step. The target groups are not homogeneous and can differ with respect to their attitudes towards the slurs and to their willingness to identify with them. This is, again, noticeable in the histories of ‘queer’ in which different reclamatory meanings collide (the confrontational ‘queer’ vs. the umbrella term ‘queer’) and of the n-word in which many members of the target group categorically oppose to the reclamatory efforts. The idea that by the fourth stage there can be multiple local reclamatory meanings of a slur can be explained by using Anderson’s (2018) employment of the notion of *communities of practice*<sup>8</sup> into analyzing slur uses. There can be many communities of practice within groups such as “African Americans” or “American LGBT activists” and therefore there can be many local reclamatory meanings of the n-word or of ‘queer’. In the case of the n-word, the initial reclamatory meaning that developed in the enslaved community (see Rahman 2012) was local—enslavers and other white people were not aware of this meaning. As the reclamatory use of the n-word became more widespread, the mainstream started to acknowledge the other meaning which lead to polysemy—one global meaning of the n-word is derogatory, and the other global meaning is the reclaimed, positive one. In the case of more than one meaning introduced in the third stage, it is possible that during the fourth stage one of them will supplant the others or that more than one meaning introduced by in-group self- or-group identification will move onto the fifth stage.

It is only in the fifth step that the reclamatory meanings of the slur enter the mainstream. This does not mean that no out-groups have heard

---

<sup>8</sup> Anderson (2018) cites Eckert and McConnell-Ginet’s (1992) definition of communities of practice: “An aggregate of people who come together around mutual engagement in an endeavor. Ways of doing things, ways of talking, beliefs, values, power relations—in short, practices—emerge in the course of this mutual endeavor. As a social construct, a community of practice is different from the traditional community, primarily because it is defined simultaneously by its membership and by the practice in which that membership engages.” (1992, 464)

the new use or even understood the new meaning of the slur. During the fifth stage of slur reclamation the new, reclaimed meaning of the slur becomes widely recognized by the public. The new meaning might become recognized but not adopted by the out-groups, as in the case of the n-word; it can become recognized and adopted by the out-groups, as in the case of 'gay'; or it can become adopted and therefore transformed because of being no longer a revolutionary term, as in the cases of 'queer' and 'Black'. It is simply impossible for the widely accepted terms to be revolutionary. In the case of more than one meaning entering the mainstream, which one of them supplants the others depends on the uptakes and the power balance—e.g., the term 'queer' as an umbrella term is less controversial and safer for the heterosexual majority than its confrontational meaning. Herbert's (2015) analysis of the risks of attempting to reclaim slurs can illuminate the process in which a revolutionary local meaning of a reclaimed slur loses its revolutionary connotation when entering the mainstream, as in the cases of 'queer' and 'Black'. For the mainstream (e.g., white people or cisgender, heterosexual people), the revolutionary meaning of the reclaimed slur is a threat to their privileged social position. The mainstream can accept a neutral, descriptive meaning of a reclaimed slur but not the revolutionary, subversive, and powerful meaning that is aimed at changing the oppressive social norms which put the targeted group in a worse social position. Herbert focuses on the cases in which a speaker attempts to use a slur in a reclamatory way and yet the audience does not recognize this speech act as reclamatory but as a standard derogatory use of a slur. Here is how she explains the negative consequences of failed attempts at reclamation: "The way the act is taken up determines the force of the act, even when this force is contrary from the original intent of the speaker. When attempts at reclamation fail, context and convention lead a hearer to give uptake to the speech act as deploying a traditional use of the slur. The force of this traditional use is to validate and re-entrench the very norms the act was intended to subvert." (Herbert 2015, 32). While I do agree with this description of reclamatory speech act failure, I believe that even more often the audience (especially people that would characterize themselves as "allies") recognizes that the speaker does not use the slur in the traditional derogatory way but fails to recognize the revolutionary nature of the speech act and the positive



evaluation encoded by the reclamatory use. The uptake distorts the intended subversive speech act<sup>9</sup> into a neutral one, and with widespread distortive uptakes and imitation the slur's reclaimed meaning ends up descriptive and hence nonthreatening to the out-groups.<sup>10</sup>

In the last stage we obtain the end result of the reclamation process which can either be polysemy or replacement. In the polysemy end result, the slur has both the derogatory meaning it had in the first stage and the reclaimed meaning recognized by the public, as in the case of the n-word. In the replacement end result, the reclaimed meaning recognized by the public replaces the derogatory one, as in the cases of 'Black' and 'gay'.

## 6. Conclusion

In this paper, I set out to investigate slur reclamation. Jeshion's (2020) Polysemy view and Bianchi's (2014) Echoic view were discussed and the histories of the reclamation of 'queer' and of the n-word were presented. I argued for incorporating the Polysemy and Echoic views into a Combined view which explains the process of slur reclamation and accounts for the examples of histories of reclaimed slurs. The Combined view accounts for the meaning change of the slur during the process of its reclamation but does not ignore the pragmatic step necessary for the introduction of a new linguistic convention. It also explains why the particular processes of slur reclamation vary with respect to both the initial intent behind the reclamatory acts and the end result of the reclamation.

---

<sup>9</sup> Here my notion of "distorting a speech act" means that the uptake differs from what the speaker intended to do with the speech act. Following Kukla (2014), I take the uptake to be the determinant of what sort of speech act has been made—if the speaker says 'Close the door!' intending to issue an order but the audience takes their speech act to be a request, the speech act in question *is* in fact a request.

<sup>10</sup> A similar process, albeit one that does not influence the local meaning of a reclaimed slur, happens when the out-group audience mistakes the in-group speaker's reclamatory use of a slur as a permission to imitate and use it. The audience recognizes the speech act as reclamatory but fails to recognize that it is prohibited that they use it.

I believe that the Combined view can be used to investigate meaning change in general, and in particular group identity-labels, differing from slurs, originating in internet slang (such as ‘incels’) and the acts of eliciting linguistic change rooted in the fight for equality (such as deeming offensive terms inappropriate and proposing new ones, as in replacing ‘retarded’ with ‘intellectually disabled’, or feminist language reforms). The latter, sometimes called *Ameliorative Projects* (see Ritchie 2021), share many characteristics with the process of slur reclamation, but what distinguishes reclamation is the echoing of the derogatory content of a slur in the second stage and the self- and group-identification with the echo in the third stage.

What is yet to be done is a detailed examination of various ongoing, finished or faded processes of slur reclamation, by means of linguistic analysis and experimental work, in order to further test the applicability of the Combined view here proposed.

### References

- Anderson, Luvell. 2018. “Calling, Addressing, and Appropriation.” In *Bad Words: Philosophical Perspectives on Slurs*, edited by David Sosa, 6–28. Oxford: Oxford University Press. [10.1093/oso/9780198758655.003.0002](https://doi.org/10.1093/oso/9780198758655.003.0002)
- Bianchi, Claudia. 2014. “Slurs and Appropriation: An Echoic Account.” *Journal of Pragmatics* 66: 35–44. <https://doi.org/10.1016/j.pragma.2014.02.009>
- Brontsema, Robin. 2004. “A Queer Revolution: Reconceptualizing the Debate Over Linguistic Reclamation.” *Colorado Research in Linguistics* 17 (1): 1–17. <https://doi.org/10.25810/dky3-zq57>
- Cepollaro, Bianca. 2020. *Slurs and Thick Terms. When Language Encodes Values*. Maryland: Lexington Books.
- Chen, Melinda Yuen-Ching. 1998. “‘I am an Animal!’: Lexical Reappropriation, Performativity, and Queer.” In *Engendering Communication: Proceedings from the Fifth Berkeley Women and Language Conference*, edited by Suzanne Wertheim, Ashlee C. Bailey, and Monica Corston-Oliver, 128–40. Berkeley, CA: Berkely Women and Language Group.
- Eckert, Penelope, and Sally McConnell-Ginet. 1992. “Think Practically and Look Locally: Language and Gender as Community-Based Practice.” *Annual Review of Anthropology* 21 (1): 461–88. [10.1146/annurev.an.21.100192.002333](https://doi.org/10.1146/annurev.an.21.100192.002333)
- Herbert, Cassie. 2015. “Precarious Projects: The Performative Structure of Reclamation.” *Language Sciences* 52: 131–8. <https://doi.org/10.1016/j.langsci.2015.05.002>

- 
- Jeshion, Robin. 2020. "Pride and Prejudiced: on the Reclamation of Slurs." *Grazer Philosophische Studien* 97 (1): 106–37. <https://doi.org/10.1163/18756735-09701007>
- Kukla, Quill. 2014. "Performative Force, Convention, and Discursive Injustice." *Hypatia* 29 (2): 440–57. <https://doi.org/10.1111/j.1527-2001.2012.01316.x>
- Rahman, Jacquelyn. 2012. "The N Word: Its History and Use in the African American Community." *Journal of English Linguistics* 40 (2): 137–71. <https://doi.org/10.1177/0075424211414807>
- Rand, Erin J. 2014. *Reclaiming Queer: Activist and Academic Rhetorics of Resistance*. University of Alabama Press.
- Ritchie, Katherine. 2021. "Essentializing Language and the Prospects for Ameliorative Projects." *Ethics* 131 (3): 460–88. <https://doi.org/10.1086/712576>
- Sperber, Dan, and Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.

## Beyond the Conversation: The Pervasive Danger of Slurs

Alba Moreno\* – Eduardo Pérez-Navarro\*\*


Received: 3 December 2020 / 1<sup>st</sup> Revised: 14 April 2021 /  
2<sup>nd</sup> Revised: 31 May 2021 / Accepted: 9 August 2021

*Abstract:* Although slurs are conventionally defined as derogatory words, it has been widely noted that not all of their occurrences are derogatory. This may lead us to think that there are “innocent” occurrences of slurs, i.e., occurrences of slurs that are not harmful in any sense. The aim of this paper is to challenge this assumption. Our thesis is that slurs are always potentially harmful, even if some of their occurrences are nonderogatory. Our argument is the following. Derogatory occurrences of slurs are not characterized by their sharing any specific linguistic form; instead, they are those that take place in what we call uncontrolled contexts, that is, contexts in which we do not have enough knowledge of our audience to predict what the uptake of the utterance will be. Slurs uttered in controlled contexts, by contrast, may lack derogatory character. However, although the kind of context at which the utterance of a slur takes place can make it nonderogatory, it cannot completely deprive it of its harmful potential. Utterances of

---

\* Universidad de Granada


 <https://orcid.org/0000-0002-6883-4357>

 Facultad de Psicología, Despacho 256, Campus Universitario de Cartuja, s/n  
18011, Granada, Spain.

 [almorenozurita@gmail.com](mailto:almorenozurita@gmail.com)

\*\* Universidad de Granada

 <https://orcid.org/0000-0002-2240-2380>

 Facultad de Psicología, Despacho 256, Campus Universitario de Cartuja, s/n  
18011, Granada, Spain.

 [edperez@ugr.es](mailto:edperez@ugr.es)



slurs in controlled contexts still contribute to normalizing their utterances in uncontrolled contexts, which makes nonderogatory occurrences of slurs potentially harmful too.

*Keywords:* Context; derogation; nonderogatory occurrences of slurs; normalization; slurs.

## 1. Introduction

Although slurs are conventionally defined as derogatory words, it has been widely noted that not all of their occurrences are derogatory. Cases of *mention*, in which we talk about the word rather than applying it to anybody, are the ones that most straightforwardly come to mind (see e.g. Hornsby 2001, 129–30). However, some full-fledged *uses* of slurs are standardly taken to be nonderogatory too. Among these, two kinds of uses have been most discussed. On the one hand, members of the target group can *appropriate* a slur in order to demarcate the group or foster solidarity or feelings of belonging, thus being able to use it in a nonderogatory way (see Bianchi 2014; Cepollaro 2017). But, on the other hand, we can also find nonderogatory uses of slurs that are not instances of appropriation—what have been called *nonderogatory, nonappropriated* (NDNA) uses of slurs (Hom 2008; see also Croom 2011, and section 2 of this paper for examples).<sup>1</sup> The fact that not all occurrences of slurs are derogatory, as mentions, appropriated and NDNA uses seem to prove, may lead us to think that there are “innocent” occurrences of slurs, i.e., occurrences of slurs that are not harmful in any sense.

The aim of this paper is to challenge this assumption. Our thesis is that slurs are always potentially harmful, even if some of their occurrences are nonderogatory. This does not mean that we take ourselves to be *prohibitionists* (see Anderson and Lepore 2013a, 2013b; see also Cepollaro, Sulpizio and Bianchi 2019, 33). That is, we do not think that it should be morally forbidden to utter a slur even when it is mentioned, for instance, for

---

<sup>1</sup> Other nonderogatory uses of slurs that have been discussed in the literature are referential (Anderson 2018) and identificatory (Zeman 2021) uses. We will briefly turn to these in section 3.

pedagogical purposes. But we do think that the utterance of a slur always comes at a moral cost, which in cases like this may be worth paying.

Our argument is the following. Derogatory occurrences of slurs (which, following Hom (2010), we call “orthodox occurrences”) are not characterized by their sharing any specific linguistic form; instead, they are those that take place in what we call *uncontrolled contexts*, that is, contexts in which we do not have enough knowledge of our audience to predict what the uptake of the utterance will be. Slurs uttered in *controlled* contexts, by contrast, may lack derogatory character. However, although the kind of context at which the utterance of a slur takes place can make it nonderogatory, it cannot completely deprive it of its harmful potential. Utterances of slurs in controlled contexts still contribute to *normalizing* their utterances in uncontrolled contexts, which makes nonorthodox occurrences of slurs potentially harmful too. It is not one of the aims of this paper to establish what makes utterances of slurs in uncontrolled contexts derogatory, nor in what sense exactly they are harmful. We just assume the common intuition that most occurrences of slurs are derogatory and in consequence harmful, and suggest that these coincide with those that take place in uncontrolled contexts. Our argument should be read as the conditional one that, *if* these occurrences of slurs are harmful, *then* all of them are potentially so.

Insofar as one of the outcomes of our work concerns the moral permissibility of mentioning a slur, it points in the same direction as Herbert (ms.), who argues that we should be careful even when merely *talking about* slurs. Her argument is that, just by mentioning these words, we already trigger harmful implicit associations.<sup>2</sup> Although we share Herbert’s concerns and reach a conclusion similar to hers, there are some differences between her work and ours that are worth commenting on. We will do so after presenting our argument.

The structure of the paper is as follows. In section 2, we review the distinction between derogatory and nonderogatory occurrences of slurs in terms of Hom’s (2010) distinction between orthodox and nonorthodox occurrences of slurs. We argue that the difference between these two kinds of occurrences does not lie in the linguistic form of the sentence uttered, but

---

<sup>2</sup> We are greatly indebted to Cassie Herbert for kindly sharing her manuscript with us.

in the context at which each of them takes place. In section 3, we flesh out what exactly this supposes by distinguishing between controlled and uncontrolled contexts; occurrences of slurs in uncontrolled contexts are always derogatory, while occurrences in controlled contexts may not be so. In section 4, we argue that occurrences of slurs in controlled contexts, even if sometimes not derogatory, have a normalizing potential that makes occurrences of slurs in uncontrolled contexts more likely. Hence, that an occurrence of a slur is nonderogatory does not mean that it does not have the potential to harm. We end the section by comparing our view with Herbert's (ms.). In section 5, finally, we discuss some of the consequences that our point may have for philosophical practice.

## 2. Nonorthodox occurrences of slurs

In this section, we survey the different cases in which occurrences of slurs have been said to be nonderogatory. These will be the cases on which we will focus in subsequent sections to discuss whether the fact that they are not derogatory means that they are not problematic in any sense. At the end of the section, we will argue that these cases include *uses* of slurs, despite attempts to reduce all nonderogatory occurrences of slurs to cases of *mention*.

According to Hom (2010, 168–69), some occurrences of slurs (which he calls “orthodox occurrences”) are nondisplaceable, while others (which he calls “nonorthodox occurrences”) are displaceable. Orthodox occurrences are nondisplaceable because they are derogatory even when embedded, while nonorthodox occurrences are displaceable in the sense that they are not always derogatory. In this paper, we do not embrace a particularly precise conception of derogation. It should be enough to say that derogation is the application to an individual of a negative moral evaluation (Hom and May 2013, 310), which is “an objective feature of the semantic contents of pejorative terms”<sup>3</sup> (Hom 2012, 397). This distinguishes derogation from

---

<sup>3</sup> This is not incompatible with accepting that some occurrences of slurs are derogatory and others are not. An occurrence of a slur can be nonderogatory because its

mere offense, which is a psychological phenomenon depending on the beliefs and values of participants in the conversation. Moreover, slurs, are opposed to mere insults, derogate its target in virtue of their belonging to a certain social group. We will assume that this kind of derogation is harmful in some way—be it because it subordinates (Kukla 2018) or dehumanizes (Jeshion 2018) its target. Let ‘*S*’ be a slur, and let us substitute it for the word that Hom uses in his examples of orthodox occurrences of slurs:

- (1) If there are *Ss* in the building, then *X* will be relieved.
- (2) There are no *Ss* in the building.
- (3) Are there *Ss* in the building?
- (4) John said that there are *Ss* in the building.
- (5) John said: ‘There are *Ss* in the building.’
- (6) In the novel, there are *Ss* in the building.

Hom takes ‘*S*’ to be derogatory in (1–6).<sup>4</sup> He takes it to be so even if it appears in the antecedent of a conditional in (1), embedded under negation in (2), as part of a question in (3), reported in indirect style in (4) and in direct style in (5), and embedded under an “in the fiction” operator in (6) (see also Hornsby 2001, 129–130; Potts 2007, 166; Croom 2011, 347; Hom 2012, 384–385; Anderson and Lepore 2013a, 30; Croom 2014, 228). However, he does not take ‘*S*’ to be derogatory in his example of a nonorthodox occurrence of a slur, which he takes from (Hom 2008, 429) and we reproduce here substituting ‘*N*’ for the (alleged) neutral counterpart of ‘*S*’:<sup>5</sup>

- (7) Institutions that treat *Ns* as *Ss* are morally deprived.

---

semantic content does not result in derogation when it interacts with features of the particular linguistic environment or context of utterance.

<sup>4</sup> As Hom (2010, n. 17) acknowledges, occurrences of ‘*S*’ such as the one in (5) are not incontrovertibly derogatory. To support the claim that they are, Hom argues that a speaker who is not a member of the target group and is not racist would be reluctant to utter (5) in front of a member of the target group, and that this is at least partly explained by the fact that the occurrence of the slur in it is derogatory. This argument seems sound enough to us.

<sup>5</sup> We do not entirely agree with the idea that slurs have neutral counterparts, but we will assume throughout this paper that they do. For discussion on this issue, see Mühlebach (2019).



Hom takes (7) to contain a nonderogatory occurrence of ‘*S*’. Note that, in this case, ‘*S*’ is not embedded in any of the ways depicted in (1–6). In this paper, we will assume that Hom’s classification coincides with the intuitions of most speakers. Thus, we will take Hom’s diagnosis that (1–6) contain derogatory occurrences of ‘*S*’ and (7) does not as part of our data.

A couple of categories can be distinguished within nonorthodox occurrences of slurs. One of the cases that most readily come to mind is that of appropriated uses of slurs. Appropriated uses of slurs are those that take place when speakers belonging to the target group aim at demarcating the group or fostering solidarity or feelings of belonging (see Bianchi 2014; Cepollaro 2017; Anderson 2018). Speakers who have appropriated a slur can use it to refer to themselves or other members of the group without derogating anyone.

But we can also find occurrences of slurs that are nonorthodox without being instances of appropriation. These have been aptly labeled “nonderogatory, nonappropriated” (NDNA) uses of slurs, an umbrella term for all nonorthodox uses of slurs that are not appropriated (Hom 2008; see also Croom 2011). An example of an NDNA use of a slur (given by Hom 2008, 429) is:

(8) There are lots of *Ns* at *Y*, but no *Ss*.

(7) would be another example of an NDNA use of a slur.

Appropriated and NDNA uses are both *uses* of slurs, but paradigmatic nonorthodox occurrences of slurs are *mentions* of them. In fact, Hornsby apparently endorses the idea that all nonorthodox occurrences of slurs are at the end of the day cases of mention:

Certainly there are occurrences of derogatory words that are utterly inoffensive: “He is not [an *S*]” can be said in order to reject the derogatory “[*S*]”; one can convey that “[*S*]” is not something one calls anyone by saying “There aren’t any [*Ss*].” But these examples do not count against their uselessness as I mean this, because they are examples in which it is part of the speaker’s message that she has no use for the word “[*S*]”. We might gloss the two sentences so that the word is mentioned rather than used: “[“[*S*]” is not what he ought to be called]; “[“[*S*]” has no application.” (Hornsby 2001, 129)

Hornsby seems to reduce all cases of nonorthodox occurrences of slurs to cases of mention rather than use. A plausible paraphrase of (7) in which ‘*S*’ is mentioned rather than used would be this:

- (9) Institutions that treat *Ns* as deserving to be called ‘*Ss*’ are morally depraved.

However, we do not think that Hom’s distinction between orthodox and nonorthodox occurrences of slurs should coincide with the distinction between use and mention, so that every sentence featuring a nonorthodox occurrence of a slur can be paraphrased as a case of mention. In fact, we do not think that the distinction between orthodox and nonorthodox occurrences of slurs is a *linguistic* distinction, in the sense that a criterion to distinguish the latter from the former can be given just in terms of the form of the sentence used. A sentence in which a slur is used can be derogatory or nonderogatory independently of whether the slur appears in the antecedent of a conditional, embedded under negation, or as part of a question, and the same happens when the slur is merely mentioned.

The relevant factor when distinguishing between orthodox and nonorthodox occurrences of slurs is the context. (1) and (7) are both cases of use, but (1) is derogatory and (7) is not. (5) and (9) are both cases of mention, but (5) is derogatory and (9) is not. The differences lie in the kind of context that we most plausibly associate with each sentence: (1) is most easily imagined as uttered in a context in which an act of derogation takes place, while (7) tends to make us picture a context in which the speaker is in fact denouncing derogatory practices, and something parallel to this can be said about (5) and (9).

Thus, the difference between orthodox and nonorthodox occurrences of slurs does not lie in the form of the sentence used, but in the context in which they take place. In the next section, we flesh out what exactly distinguishes some contexts from others. However, as we will see, the difference has a limited impact, since it makes nonorthodox occurrences of slurs less dangerous, but not strictly not dangerous.

### 3. Controlled and uncontrolled contexts

The upshot of the previous section was that whether an occurrence of a slur is orthodox or nonorthodox depends on the context at which it takes place. In this section, we take a closer look at the kinds of contexts that make an occurrence of a slur orthodox or nonorthodox. In particular, we identify nonorthodox occurrences of slurs with those that can take place in what we call “controlled contexts” and orthodox occurrences of slurs with those that take place in “uncontrolled contexts”. However, we will see in section 4 that, even if the distinction between controlled and uncontrolled contexts can help us rank occurrences of slurs according to their derogatory character, part of slurs’ power to cause harm is distributed equally across the categories distinguished here.

Communication is a risky business. There are a number of factors that can have an impact on the kind of effect that a given utterance will have, and most of them escape our control. When communicating, we often have to manage without knowing what our audience knows or what their expectations are. Still, even if rare, contexts can be found in which we can predict with reasonable accuracy what the consequences of a given utterance will be. We call these “controlled contexts”.<sup>6</sup> When we are talking about utterances including a slur, an example of a controlled context would ideally be a pedagogic one, and another, more contentious one would be that in which a slur is successfully used in an ironic way.

Here is an example of a pedagogic occurrence of a slur. Our son Dani comes home from school and says his friend *Y* says his other friend *X* is an *S*. Later on, we tell Dani he should never say that word again. ‘What word?’, he says. He has not forgotten it, but honestly cannot recall which one of the words he has pronounced we are forbidding him from saying. We feel forced to pronounce ‘*S*’ in order to make sure he knows what term we are referring to, so we do—we say ‘We don’t call people ‘*S*’, that’s an ugly thing to say.’ We have uttered a slur, even if we have only mentioned

---

<sup>6</sup> Of course, whether a given context is a controlled one will in many cases not be a settled matter. We leave for further work to offer precise criteria for a context to fall under this label.

it.<sup>7</sup> But we had no other option, and we can be sure that by doing this we have not insulted anyone—if anything, we have prevented Dani from insulting anyone, even if from unintentionally doing so. We know enough about our own son to guarantee that he has understood that we were not insulting anyone. Here, the occurrence of ‘*S*’ is nonderogatory.

Here is another kind of case in which we can say that the utterance of a slur has taken place in a controlled context. This time, we are not talking about a mere mention of a slur, but about a full-blown use—an ironic use. We are a progressive group of friends who would never as much as mention a slur in front of strangers, much less use it to insult a person on grounds of her belonging to a given group. However, we find fun in imitating bigots’ mannerisms, and enjoy inner jokes that include ironic uses of ‘*S*’. We are completely sure that all our friends in the group share our sensibility, and that none of them will take us to aim at insulting anyone. We think it is intuitive to take occurrences of slurs such as these to be nonderogatory, whatever the form of the sentences in which they appear.

Other nonderogatory uses of slurs that have recently been described are referential (Anderson 2018) and identificatory (Zeman 2021) uses. Referential uses take place when members of the target group use a slur to address other members without any intention to appropriate the term, while identificatory uses take place when they simply take the word to be the one that refers to the group they take themselves to belong to. We take these uses to take place in controlled contexts too, as the speaker’s group membership is salient enough for her to be confident that the audience will understand that she did not mean to insult, just like happened in Dani’s case.

If controlled contexts are those in which we can be sure about the other participants’ knowledge and expectations, almost all contexts in which we can find ourselves are uncontrolled ones. It is difficult to know anyone as

---

<sup>7</sup> It could be argued that pedagogic contexts not only allow for nonderogatory *mentions* of slurs, but also for nonderogatory *uses* of them. For instance, someone might claim, we could have also said to Dani ‘There are no *S*s, only *N*s.’. However, we find it hard to see this sentence as nonderogatory—the fact that the speaker has seen the need to categorize Dani’s friend as belonging to the target group, even if in a supposedly neutral way, makes the sentence problematic. See again Mühlebach (2019).

well as we know our own children or our closest friends, and in many cases we hardly have any relevant information about our audience. Consider our daily interactions with strangers, and the limiting case of the completely uncontrolled context in which public communication takes place. When we utter a slur in an uncontrolled context, our audience has every reason to attribute to us a negative attitude toward a given group, and we cannot reasonably expect not to be attributed such an attitude, which is what, in an intuitive sense, means to derogate (see section 2). Thus, in uncontrolled contexts, which are most of the contexts, occurrences of slurs are derogatory.

As advanced before, it lies beyond the scope of this paper to offer an explanation of exactly how slurs derogate when uttered in an uncontrolled context. There are a number of proposals in the market that aim at accounting for this fact. Some of these views rely on specifically derogatory content that is part of what is said (Hom 2008, 2010, 2012; Hom and May 2013), conventionally implicated (Potts 2007; Copp 2009; Williamson 2009; McCready 2010; Whiting 2013), or presupposed (Macià 2002; Schlenker 2007; Cepollaro and Stojanovic 2016; Marques and García-Carpintero 2020). Some of them, by contrast, explain the derogatory character of slurs without appealing to specifically derogatory content (Anderson and Lepore 2013a, 2013b). At any rate, these are all different ways of accounting for the widely held intuition that, in most of the cases, occurrences of slurs are derogatory.

A plausible objection is that slurs can occur in a derogatory way in controlled contexts too. I may know exactly what the reaction of the audience to my utterance of a slur will be, and know this reaction to be one that will precisely result in derogation. In this case, the occurrence of the slur will be derogatory even if it takes place in a controlled context. Note, however, that what distinguishes controlled and uncontrolled contexts is that occurrences of slurs in the former *can* be nonderogatory, not that they will always be so.

Another plausible objection, mirroring the one above, is the following. In uncontrolled contexts, we cannot be sure that the uptake of our utterance will fail to derogate, but this does not mean that it will derogate. It may happen that, just by chance, every single member of the audience

understands the occurrence of the slur as nonderogatory, even if we are not able to predict that this will be the case. For instance, all bystanders who hear Dani utter a slur could assume that he does not know what the word means.<sup>8</sup> Note, however, that we have characterized uncontrolled contexts as those in which hearers have *every reason* to attribute to us a negative attitude toward the target group, which we cannot in turn *reasonably* expect not to be attributed to us. In this case, the audience can refuse to attribute the negative attitude to us. However, inasmuch as they would be warranted in so doing, our utterance can be taken to be derogatory.

Of course, what counts as good reason is a highly context-dependent issue, and some contexts might make it reasonable not to attribute a negative attitude to the speaker. This may be the case, for instance, when the audience knows that the speaker is a decent person. Note, however, that this will only be warranted if the audience knows not only that the speaker is a decent person, but also that the speaker is aware that this is publicly known. In this case, the audience will have reason to believe that the utterance is not derogatory, but we will no longer be facing an uncontrolled context. If the relevant piece of public knowledge is missing, as should happen in an uncontrolled context, and the speaker still chooses to utter the slur, the audience can legitimately conclude that she is comfortable with being attributed a negative attitude.

This idea that, in uncontrolled contexts, it is reasonable to attribute the utterer of a slur a negative attitude toward the target group no matter what her actual attitudes are is similar to one that has been defended by Lasersohn (2007). This idea is a key component in Lasersohn's explanation of the hyperprojectivity of slurs' derogatory content. Hyperprojectivity is the phenomenon whereby the derogatory content of slurs is able, in many occasions, to survive in grammatical constructions that would usually block presuppositional content. Lasersohn defends that this fact is compatible with a presuppositional account of slurs by providing the following explanation. According to Lasersohn, slurs are emotionally charged terms, so uttering them entails a social risk. Lasersohn believes that speakers are aware of the social burden of slurs, and this is the main reason why most

---

<sup>8</sup> Thanks to an anonymous reviewer for *Organon F* for suggesting this objection to us.

speakers avoid uttering slurs—because, whatever their particular attitudes, they are aware that they can reasonably be attributed bigotry. Precisely because of this, when a speaker does utter a slur, it makes sense to think that she is comfortable with being identified as a bigot, and this is how the derogatory content of the slur projects where most presupposed content does not (Lasersohn 2007, 228). Like Lasersohn, we think that, if a speaker utters a slur in an uncontrolled context in which it is even merely possible that someone understands the occurrence of the slur as derogatory, it makes sense to take the speaker to be comfortable with this possibility, and thus to take the occurrence to be actually derogatory.

#### 4. The normalizing potential of slurs

We have seen that slurs are derogatory in uncontrolled contexts, but not in controlled contexts such as pedagogic and ironic ones. Still, no matter how carefully we arrange the current context to make sure that the utterance of a slur does not have the kind of effect we want to avoid, it will facilitate ulterior occurrences of the term. In particular, it will make the slur more likely to appear in uncontrolled contexts in which the utterance of the slur is derogatory. In this section, we explore how this could be the case with the two kinds of occurrences of slurs that we presented in the previous section—pedagogic mentions and ironic uses of slurs. If even the apparently most “innocent” occurrences of slurs, such as those that take place in pedagogic contexts, are potentially harmful, it is natural to conclude that *all* occurrences of slurs are potentially harmful.

Let us start with the irony case. Remember that, in this case, we are a progressive group of friends who enjoy using slurs in an ironic way to make fun of bigots. However, we have ironically used ‘*S*’ in our friend group so many times that we have deprived it of its forbidden character—it no longer makes us uncomfortable to hear the word, which makes its utterance in uncontrolled contexts more likely now. This nonderogatory use of a slur thus normalizes derogatory occurrences of the word, and is potentially harmful in this sense. Of course, our friend group could be careful enough not to let uses of slurs slip out of the controlled context. This is why ironic uses of slurs such as these are not harmful *tout court*, but potentially

harmful. But, since potential harm implies actual danger, these uses are dangerous *tout court*.

Now, take the example involving a pedagogic mention of a slur, also described in the previous section. Remember that, in this case, we feel forced to utter the word ‘*S*’ in order to make our son Dani aware that he should not call anyone an *S*. However, we have taught Dani what ‘*S*’ means, thus giving him the tools to use the word to insult if he wants to do so at some point. Note that, at least in this case, the risk that Dani grows up to use ‘*S*’ as an insult may be worth it: as we will see, preventing an actual risk may be preferable to preventing a virtual one. In this sense, this case might strike us as clearer than the previous one. The normalizing potential is similar in both cases, though. The difference is that, in the pedagogic case, it is clearer how the benefits could outweigh the potential harm. In the irony case, all we have on the positive side is the fun we have with our friends. Referential and identificatory uses of slurs are closer to pedagogic mentions than to ironic uses in this respect. Like with pedagogic mentions, however, there is still the risk that these uses facilitate ulterior occurrences of the term in contexts in which the group membership of the speaker, although salient, does not make the audience understand such occurrences as nonderogatory because the speaker does not belong to the target group. Hence, the moral here is that we may have full control over the present context, but we do not have control over all possible future contexts. Thus, slurs always have normalizing potential. The slur might not be problematic in the context at which it is uttered, but it may reveal itself to be so as we look beyond the original conversation and consider other exchanges that might be facilitated by the original utterance. We take something to be dangerous whenever it *may* cause harm, even if it does not actually do so. Insofar as occurrences of slurs are always potentially harmful, therefore, we take them to be always dangerous.

As we said in the introduction to this paper, our idea that even non-derogatory occurrences of slurs can be harmful should not suffice to classify us as *prohibitionists* (see Anderson and Lepore 2013a, 2013b; see also Cepol-  
laro, Sulpizio and Bianchi 2019, 33)—we do not think utterances of a slur should be forbidden *tout court*. We think there are some practical consequences to the categorization of some uses of slurs as appropriated or



NDNA, and that the distinction between use and mention has practical consequences when it comes to slurs too. It may be permissible to mention a slur in certain contexts, just like it may be permissible to make an appropriated or an NDNA use of a slur. This marks a difference between these occurrences of slurs and their full-blown, derogatory uses. But we should be aware that these practices come with a moral cost too. The price may be worth paying, of course. It just misrepresents our moral life to assume that it consists in choosing the only permissible thing to do in each case; rather, we assess the moral costs and benefits of each course of action, decide what weight to give to each, and act in consequence.<sup>9</sup> The moral benefits of performing a certain utterance of a slur might outweigh the pervasive moral cost we have described, and so it might be worth it to utter the slur.

As we also advanced in the introduction to this paper, the stance that merely mentioning a slur, as we do in the pedagogic case, can be reprehensible too has been defended before us. Herbert (ms.) argues that the practice of offering examples of slurs, which is widespread in philosophy, may cause harm just like using them does. To conclude this, she relies on empirical evidence found by Carnaghi and Maass (2008) and Fasoli, Paladino, Carnaghi, Jetten, Bastian and Bain (2015) that it is mere exposure to a slur, rather than specifically exposure to *uses* of a slur, that triggers negative implicit associations concerning the target group. The question that Herbert asks herself (is it morally permissible to mention a slur?) is precisely one of those we have set ourselves to answer in this paper, and the reply she offers is akin to ours—a refusal to give a context-independent answer, together with an invitation to be extremely careful when deciding whether to mention a slur. However, there are some differences between Herbert’s work and ours that we think make this paper a worthy contribution to the debate.

First, Herbert’s work focuses on mentions of slurs, while ours also covers some of their uses, such as ironic ones. Of course, this does not mean that Herbert’s point applies only to mention. The way her argument goes, mentions of slurs are potentially harmful just in virtue of their being *occurrences* of slurs, so her conclusion should apply to any occurrence of these words, including uses in general and ironic uses in particular. But, while most of

---

<sup>9</sup> For a really insightful guide on what particular factors to consider when deciding whether a slur is worth mentioning, see Herbert (ms.).

the discussion in Herbert's paper concerns mentions of slurs in academic and journalistic environments, ours concerns ironic uses and mentions of slurs in *pedagogic* contexts.

Second, our explanation of how these occurrences of slurs can end up being harmful is more general than Herbert's. While she relies on implicit associations, our argument is compatible with different proposals as to what mechanism accounts for the pernicious effects that uttering a slur may have, possibly including the appearance of implicit associations like the ones described by Herbert. If (contrarily to the evidence we now have) occurrences of slurs turned out not to elicit implicit associations *per se*, our work would still provide a schema that could be completed with an alternative mechanism.

## 5. Concluding remarks

In this paper, we have argued that the mere utterance of a slur has a certain kind of impact—it normalizes further occurrences of the word. This is so even in those cases in which the slur does not derogate anyone. Nothing prevents the slur from being used in a derogatory way in the ulterior occurrences normalized by these ones, so even nonderogatory occurrences make it more likely for derogation to take place at some point in the future.

Our proposal has consequences for philosophical practice as we know it. Our point is that mentions of slurs in academic papers are potentially harmful too: even if, not being used, they are not derogatory, they facilitate ulterior occurrences of the slur in question too. Of course, not all philosophers are comfortable with mentioning slurs even if they take it to be necessary. Rebecca Kukla, for instance, says:

By flagging that I will be mentioning slurs and reminding the reader that even the mention of slurs can harm, I hope to frame these mentions in a way that allows readers to be conscious of such effects and to try to minimize it. I also use scare quotes around the slurs throughout, to help avoid normalizing them as part of everyday speech, and in the hope of marking them at the

visual level as problematic terms that I am not uttering in my own voice and that are not to be taken for granted as readable. (Kukla 2018, 24)

We do not think the use of scare quotes blocks normalization, but again, this does not mean that Kukla's mentions of slurs are necessarily unjustified. The moral cost might be worth assuming in this case. We keep our doubts, however, that it is worth assuming in cases in which more examples than the strictly necessary are given.

We do not think that a criterion can be found by which certain occurrences of a slur should be allowed, but by the same token we do not think that there is a criterion that forbids the rest of occurrences *tout court*. Our point is that uttering a slur always comes at a moral cost, and it is the responsibility of the speaker, or the philosopher who writes a paper on slurs, to assess such cost and decide whether it is worth it to mention a word to explain to a child that it should never be used or to give one more example of a slur in a paper addressed to an audience that is assumed to know what slurs are.

### Acknowledgments

This paper has been funded by the Spanish Ministry of Science and Innovation under the research project "Disagreement in Attitudes: Normativity, Affective Polarization and Disagreement" (PID2019-109764RB-I00), by the Regional Government of Andalusia under the research projects "Public Disagreements, Affective Polarization and Immigration in Andalusia" (B-HUM-459-UGR18) and "The Inferential Identification of Propositions: A Reconsideration of Classical Dichotomies in Metaphysics, Semantics and Pragmatics" (P18-FR-2907), and by the University of Granada under a "Contrato Puente" fellowship and the excellence unit FiloLab-UGR (UCE. PPP2017.04). The authors would also like to thank Alex Davies, María José Frápolli, Andrés Soria, Neftalí Villanueva, Dan Zeman, and two anonymous reviewers for *Organon F*, as well as audiences at EvalLang-2019, FiloLab International Summer School 2019, Epistemological and Cognitive Analyses of Cognition, Beliefs and Knowledge, and the IX Meeting of the Spanish Society for Analytic Philosophy, for their helpful comments and suggestions.

## References

- Anderson, Luvell. 2018. "Calling, Addressing, and Appropriation." In *Bad Words: Philosophical Perspectives on Slurs*, edited by David Sosa, 6–28. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198758655.003.0002>
- Anderson, Luvell, and Ernie Lepore. 2013a. "Slurring Words." *Noûs* 47 (1): 25–48. <https://doi.org/10.1111/j.1468-0068.2010.00820.x>
- Anderson, Luvell. 2013b. "What Did You Call Me? Slurs as Prohibited Words." *Analytic Philosophy* 54 (3): 350–63. <https://doi.org/10.1111/phib.12023>
- Bianchi, Claudia. 2014. "Slurs and Appropriation: An Echoic Account." *Journal of Pragmatics* (66): 35–44. <https://doi.org/10.1016/j.pragma.2014.02.009>
- Carnaghi, Andrea, and Anne Maass. 2008. "Derogatory Language in Intergroup Context: Are "Gay" and "Fag" Synonymous?" In *Stereotype Dynamics: Language-Based Approaches to the Formation, Maintenance, and Transformation of Stereotypes*, edited by Yoshihisa Kashima, Klaus Fiedler, and Peter Freytag, 117–34. New York: Lawrence Erlbaum Associates.
- Cepollaro, Bianca. 2017. "Slurs as the Shortcut of Discrimination." *Rivista di Estetica* (64): 53–65. <https://doi.org/10.4000/estetica.2063>
- Cepollaro, Bianca, and Isidora Stojanovic. 2016. "Hybrid Evaluatives: In Defense of a Presuppositional Account." *Grazer Philosophische Studien* 93 (3): 458–88. <https://doi.org/10.1163/18756735-09303007>
- Cepollaro, Bianca, Simone Sulpizio, and Claudia Bianchi. 2019. "How Bad Is It to Report a Slur? An Empirical Investigation." *Journal of Pragmatics* (146): 32–42. <https://doi.org/10.1016/j.pragma.2019.03.012>
- Copp, David. 2009. "Realist-Expressivism and Conventional Implicature." *Oxford Studies in Metaethics* 4: 167–202. <https://doi.org/10.1017/S0265052500002880>
- Croom, Adam M. 2011. "Slurs." *Language Sciences* 33 (3): 343–58. <https://doi.org/10.1016/j.langsci.2010.11.005>
- Fasoli, Fabio, Maria Paola Paladino, Andrea Carnaghi, Jolanda Jetten, Brock Bastian, and Paul G. Bain. 2015. "Not "Just Words": Exposure to Homophobic Epithets Leads to Dehumanizing and Physical Distancing from Gay Men." *European Journal of Social Psychology* 46 (2): 237–48. <https://doi.org/10.1002/ejsp.2148>
- Herbert, Cassie. ms. "Talking About Slurs."
- Hom, Christopher. 2008. "The Semantics of Racial Epithets." *The Journal of Philosophy* 105 (8): 416–40. <https://doi.org/10.2307/20620116>
- Hom, Christopher. 2010. "Pejoratives." *Philosophy Compass* 5 (2): 164–85. <https://doi.org/https://doi.org/10.1111/j.1747-9991.2009.00274.x>
- Hom, Christopher. 2012. "A Puzzle About Pejoratives." *Philosophical Studies* 159 (3): 383–405. <https://doi.org/10.1007/s11098-011-9749-7>

- Hom, Christopher, and Robert May. 2013. "Moral and Semantic Innocence." *Analytic Philosophy* 54 (3): 293–313.  
<https://doi.org/https://doi.org/10.1111/phib.12020>
- Hornsby, Jennifer. 2001. "Meaning and Uselessness: How to Think About Derogatory Words." *Midwest Studies in Philosophy* 25 (1): 128–41.  
<https://doi.org/10.1111/1475-4975.00042>
- Kukla, Rebecca. 2018. "Slurs, Interpellation, and Ideology." *The Southern Journal of Philosophy* 56 (S1): 7–32. <https://onlinelibrary.wiley.com/doi/10.1111/sjp.12298>
- Lasersohn, Peter. 2007. "Expressives, Perspective, and Presupposition." *Theoretical Linguistics* 33 (2): 223–30. <https://doi.org/10.1515/TL.2007.015>
- Macià, Josep. 2002. "Presuposición y significado expresivo." *Theoria: Revista de Teoría, Historia y Fundamentos de la Ciencia* 3 (45): 499–513.
- Marques, Teresa, and Manuel García-Carpintero. 2020. "Really Expressive Presuppositions and How to Block Them." *Grazer Philosophische Studien* 97 (1): 138–58. <https://doi.org/10.1163/18756735-09701008>
- McCready, Elin. 2010. "Varieties of Conventional Implicature." *Semantics and Pragmatics* 3 (8): 1–57. <https://doi.org/10.3765/sp.3.8>
- Mühlebach, Deborah. 2019. "Semantic Contestations and the Meaning of Politically Significant Terms." *Inquiry*.  
<https://doi.org/10.1080/0020174X.2019.1592702>
- Potts, Christopher. 2007. "The Expressive Dimension." *Theoretical Linguistics* 33 (2): 165–98. <https://doi.org/10.1515/TL.2007.011>
- Schlenker, Philippe. 2007. "Expressive Presuppositions." *Theoretical Linguistics* 33 (2): 237–45. <https://doi.org/10.1515/TL.2007.017>
- Whiting, Daniel. 2013. "It's Not What You Said, It's the Way You Said It: Slurs and Conventional Implicatures." *Analytic Philosophy* 54 (3): 364–77.  
<https://doi.org/10.1111/phib.12024>
- Williamson, Timothy. 2009. "Reference, Inference, and the Semantics of Pejoratives." In *The Philosophy of David Kaplan*, edited by Joseph Almog, and Paolo Leonardi, 137–59. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195367881.003.0009>
- Zeman, Dan. 2021. "A Rich-Lexicon Theory of Slurs and Their Uses." *Inquiry*.  
<https://doi.org/10.1080/0020174X.2021.1903552>

## Unmentionables: Some Remarks on Taboo

Stefano Predelli\*

Received: 20 October 2020 / Accepted: 10 May 2021

*Abstract:* This paper discusses the phenomenon of linguistic taboo. It contrasts that phenomenon with the truth-conditional and non-truth-conditional dimensions of meaning, paying particular attention to slurs and coarseness. It then highlights the peculiarities of taboo and its meta-semantic repercussions: taboo is a meaning-related feature that is nevertheless directly associated with the tokening process. In the conclusion, it gestures to the role of taboo within a theory of linguistic action and the standard framework for conversational exchanges. On these results, I am going to end by looking at some of the harms that epistemic injustice inflicts upon its victims.

*Keywords:* Taboo; non-truth-conditional meaning; derogation; coarseness; register; David Kaplan.


Here is a (vague) truism: our utterances are subject to a type of normative evaluation independent from the information they encode. In what follows, I discuss this idea by focusing on a peculiar phenomenon, linguistic *taboo*.

There are several reasons why taboo, in my sense of that term, is a theoretically interesting affair. For one thing, it seems partially related with

---

\* University of Nottingham

 <https://orcid.org/0000-0002-8375-4611>

 Department of Philosophy, Room C43 Humanities Building, NG7 2RD Nottingham, United Kingdom

 [stefano.predelli@nottingham.ac.uk](mailto:stefano.predelli@nottingham.ac.uk)



more widely studied realizations of so-called non-truth-conditional meaning, as in the cases of register or derogation mentioned throughout this essay. Indeed, at least among us, the most likely candidates for anything in the vicinity of taboo are certain coarse expressions, and/or certain instances of racial or ethnic slurs. Yet, as I explain in what follows, taboo maintains a distinctive dimension: the conventional regularities responsible for the taboo status of an expression occupy an idiosyncratic niche, separate both from the classic treatments of truth-conditions and from the more recent semantic frameworks for expressives, honorifics, and all that goes with them.

To anticipate, taboo expressions are *unmentionable* expressions: what warrants a negative reaction to taboo words is their sheer display, that is, their mere occurrence. And so, their charged status derives from the presence of their tokens, including realizations within quotation marks or merely accidental occurrences. If, as it seems plausible, the taboo status of an expression is part and parcel of its conventional profile, it follows that the conventions in question must belong to areas of inquiry other than those traditionally covered by semantics – at least on a standard understanding of semantics as the study of conventional meaning, in its truth-conditional and non-truth-conditional guises.

In section one, I swiftly sketch certain features of current semantic theorizing, with particular attention to the non-truth-conditional domain. I do so for two reasons: as a preliminary background for the aforementioned partial relationships between these phenomena and taboo, but also as a term of contrast intended to highlight taboo's most intriguing idiosyncrasies. I propose a preliminary analysis of the peculiarly unmentionable status of taboo words in sections two and three, where I focus on the relationships between taboo, pure quotation, and related phenomena. Section four ventures a positive hypothesis: some aspects of the conventional profile of an expression, first and foremost its taboo status, are properties that belong in the province of the theory of action, rather than of semantics. Section five wraps things up with a few tentative remarks on the characteristic effects of the actions in question, taking as my starting point the idiosyncratic role of taboo within the study of conversational exchanges.

## 1. Non-truth-conditional meaning

One's utterances can be subjected to a straightforward form of criticism, in the sense that what they encode happens to possess undesirable properties. So, a speaker may be chastised for uttering falsehoods, for asserting what she does not believe to be the case, or for putting forth an uninformative or irrelevant claim. Or else, she may be criticized for revealing information that should have been kept secret, or for speaking of a subject which ought not to be brought up in polite company. Yet, in other cases, the object of our normative assessment seems to be more closely related to one's mode of expression, rather than to the properties of encoded content: what has been said may well have been true, relevant, or interesting, but the way in which that content has been presented is judged to be offensive, inappropriate, or objectionable.

The *justifications* for these sorts of assessments presumably ensue from significantly different considerations. For instance, falsehood, insincerity, and irrelevance apparently invoke normative constraints other than those in place for our disdain towards profanities or offensive categorizations. Yet, at least according to widespread consensus, the *source* of these outcomes remains within the domain of semantic inquiry: *modulo* the distinction I am about to mention, both what an expression encodes and the aforementioned additional effects achieved by its use depend on the conventional properties naturally categorized as parts of its *meaning*.

The distinction that I have in mind is familiar enough: certain effects of the employment of an expression ensue from a dimension of conventional meaning that is different from the type of meaning responsible for its truth-conditional contributions. Cases of register, honorifics, and slurs are perhaps the most widely discussed exemplars in this respect. So, intuitively, my use of 'tummy' at the doctor's may well contribute to true and informative information about my health, but it remains inappropriate *qua* exemplar of so-called Child Directed Speech, that is, as an instance of a type of register unsuitable for the interaction among adults. By the same token, the description of a German national by means of 'Kraut' may well be accurate



and relevant, while remaining an appropriate target of disapprobation from the viewpoint of non-xenophobic decency.<sup>1</sup>

This intuitive divide has spurred a lively and fruitful semantic debate, focused on the peculiarity of the kind of meaning responsible for outcomes of register, coarseness, slurring, and all the rest. The buzzwords here are familiar enough and need not be rehearsed here: unlike more familiar forms of meaning, the conventions in question engender peculiar outcomes of *non-displaceability* or projection, they play a *non-at-issue* role in the economy of conversation, and/or they display peculiarly *expressive* characteristics. Accordingly, as widespread philosophical jargon puts it, conventional meaning distributes along two different dimensions: a *truth-conditional* dimension responsible for the contribution of truth-conditionally relevant content, and a *non-truth-conditional* aspect devoted to those other peculiarities in an expression's conventional profile.<sup>2</sup>

Two aspects in this research program are worthy of note. The first decrees that issues of register, derogation, and all the rest ensue from meaning. The second focuses on the type of meaning it is. It is this second issue that has received most of the attention in the current literature: non-truth-conditional meaning is meaning all right, but it is a meaning of a special sort,

---

<sup>1</sup> On register and related issues see for instance (Cruse 1986), (Allan 1990), and (Allan and Burrige 2006). For an influential essay on honorifics, see (Harada 1976); see also (Holmes 1992). For a sample of the discussion of derogatory terms, see (Potts 2003), (Hom 2008), (Richard 2008), (Williamson 2009), (McCready 2010), and (Anderson and Lepore 2013). For a study of the relationships between slurs and register, see (Diaz-Legaspe et al. 2020) and, for my own views on these matters, see (Predelli 2013).

<sup>2</sup> The labels are not entirely perspicuous. At least in some views, what falls under the non-truth-conditional side is explicable in terms of truth-evaluable content, as in the case of 'Kraut' and the content that the speaker disparages Germans. Still, even in those views, this type of information is extraneous to the truth-conditions of, say, 'Angela is a Kraut', thereby providing at least partial justification for the standard description of its source as 'non-truth-conditional'. Regarding (different versions of) this paradigm see, among many, (Kaplan 1999), (Kratzer 1999), (Potts 2003), (McCready 2010), and (Gutzmann 2015); for my own views see (Predelli 2013). For discussions of non-displaceability see in particular (Kratzer 1999), (Potts 2003), (Sauerland 2007), (Amaral et al. 2007), and (Simons et al. 2010).

worthy of being unveiled in all its multifarious manifestations. The starting point of the ensuing debates, namely the semantic dignity of the non-truth-conditional dimension, is often accepted without further ado. But here, presumably, the proof is in the pudding: a non-truth-conditional dimension of meaning may safely be taken on board, as long as it yields desirable and theoretically fruitful outcomes.

For the record, for me, the pudding is worth the effort. In particular, the non-truth-conditional dimension apparently gives rise to phenomena and regularities that mirror, *mutatis mutandis*, those ensuing from more familiar sources of meaning. For instance, ‘Angela is a Kraut but I never derogate Germans’, though possibly true, seems to engender a tension most naturally explainable in terms of the contrast between the derogation included in the meaning of ‘Kraut’, and the content encoded in the second conjunct. Or else, in Italian, ‘ti ho invitato a pranzo’ (‘I treated you [informal] to lunch’) apparently bears some close relationship to ‘the speaker is in a relation of familiarity with the addressee’, even though it does not entail it.

And so, without further ado (and without argument) I happily go along with the familiar multi-dimensional approach to conventional meaning and to semantics. For me, then, ‘stomach’ and ‘tummy’, ‘German’ and ‘Kraut’, or, to cite an example that will play some role in what follows, ‘to copulate’ and ‘to fuck’ are truth-conditionally on a par, but they are not synonymous: by virtue of their meaning, the latter are instances of, respectively, Child Directed Speech, slurring, and coarseness. I accept all of this with nonchalance because my topic is not non-truth-conditional meaning, but the contrast between the dimensions of meaning to which I have alluded thus far and a different property of certain expressions. That property is the protagonist of this essay, namely *taboo*.

## 2. Towards taboo

The last example in the foregoing paragraph was not out of place in an academic journal: coarseness, the phenomenon exhibited by ‘fuck’, is surely an appropriate instance of non-truth-conditional meaning, side by side with register, honorifics, or derogation. It is, unsurprisingly, a word that has

received the attention it deserves, including a monograph with essays directly devoted to it and to related affairs such as ‘up yours’ (Zwicky et al. 1971). In its adverbial form, it even appears in the title of an article in the academically dignified journal *Theoretical Linguistics*, unsurprisingly in an issue devoted to non-truth-conditional meaning in all of its manifestations (Geurts 2007).

Of course, neither I nor the authors of those essays inappropriately abandoned the terse register required for academic publication. In a nutshell, we all freely *mentioned* a coarse expression, but we refrained from using it. And yet, that word probably stood out in my list of specimens of the non-truth-conditional domain. Of course, only the prissiest of readers would have been inclined to chastise my display. Still, even my most blasé audience did not fail to note its colourful appearance. Even dignified linguists such as the contributors to the aforementioned collection, though officially engrossed in the linguistic properties of ‘fuck’, did not conceal the transgressive gusto with which they put that expression on the printed page.<sup>3</sup>

This phenomenon is even more apparent when what is at issue is not what most of my readers are likely to consider a minor infraction of proper decorum. Derogatory terms with a history and potency much stronger than my relatively tame ‘Kraut’ resist unfettered mention to a greater extent than instances of coarseness, and they do so on the basis of a normative stance grounded on less superficial principles. I remain confident that, at most, mention of ‘fuck’ puts a cursory smile on some readers’ lips. Other more alarming instances may in the end legitimately appear within quotation marks. Still I, and for all I know most of the students of slurs, steer away from causing offence by listing as their exemplars mild-mannered affairs such as ‘Kraut’ or, in Michael Dummett’s case, downright antiquated exemplars such as ‘Boche’ (Dummett 1973).

---

<sup>3</sup> In her discussion of slurs and offensiveness Renée Bolinger notes how “there is still something ... offensive about listing ... slurs explicitly” (Bolinger 2017, 443), and how “in some cases a speaker may rightly be censured for directly mentioning the slur” (Bolinger 2017, 451). Her topic is, of course, importantly different from mine: here, coarseness and slurs only occur as negotiable evidence of the limited forms of linguistic taboo available in our society. But the recognition of a certain ‘resistance to mention’ remains apt.

It may be argued that all of the above indulges in excessive delicacy, a puerile fascination with the sheer sight of ‘fuck’ or, in the case of those most distasteful racial slurs, yet another case of political correctness gone crazy. That may, in the end, be the case. But it is a case that needs to be made. The question whether ‘fuck’ may at all be mentioned in an academic journal may well be replied in the positive, but it remains an intelligible question, unlike, say, the question whether one may mention ‘rabbit’ or ‘Aristotle’. And the question whether this tolerance ought to extend to other instances is one that should not be taken unreflectively.

This peculiarity of the expressions under discussion, that is, their noteworthy occurrence within mention, is further testified by the existence of conventionalized locutions that designate those words without indulging in their displays. For instance, although I could refer to a common coarse designator of sexual activity by enclosing that verb within quotation marks, I could more delicately have done so by using the dedicated description ‘the f-word’. No prizes for guessing its designatum: that description is properly in the singular, and what it designates is neither ‘French’ nor ‘fries’.<sup>4</sup> Similarly, for any moderately informed contemporary speaker, no guessing is required when it comes to ‘the n-word’, a description that pursues the wonders of mention without indulging in spelling out the unmentionable.

These features of ‘fuck’ and of particularly charged racial slurs deserve further attention. They are, in a sense, conventional: the distasteful aspects ensuing from the appearance of that four-letter word are surely not natural properties associated with its sound or with the shape of its written form. And yet, these are conventional features that do not fit the semantic frame of mind with which most of us approach register, derogation, or coarseness – or, even more clearly, designation or entailment. That they do not is testified by the very phenomenon to which I have called attention: their resistance to pure quotation.

After all, in a sense, pure quotation is a device *designed* so as to absorb meaning away. And so, in using the six-character quotational term “fuck” (that is, ‘fuck’ flanked by quotation marks), you neither speak of sexual intercourse nor engage in coarse verbal behaviour. At the very least: pure

---

<sup>4</sup> See (Hughes 1991), (Harris 1987), (Davis 1989); see also (Zwicky 2003).

quotation isolates you, the speaker, from the truth-conditional and non-truth-conditional commitments associated with what occurs inside the quotes. From the viewpoint of what is being used, then, the result of appending quotation marks is not a function of any of the semantic properties of that to which the quotes are appended. And so, Quine may perhaps have exaggerated the semantic inertia achieved by pure quotation when he proposed that ‘cat’ occurs in “cat” no more interestingly than it does in ‘cat-atonic’.<sup>5</sup> Yet, semantic inertia remains the name of the game, as testified by the well-formedness of instances where what is included in quotation marks is not a well-formed expression at all, as in the unobjectionable sentence “xrt’ is not an English word’. And so, the connotations of ‘fuck’ that still reverberate once that term is merely being mentioned must lie on a plane importantly different from that appropriate for all the dimensions of meaning to which I have alluded above, be they truth-conditionally significant properties or instances of non-truth-conditional meaning.

Quine’s ‘catatonic’ is telling in this respect, since noteworthy repercussions may ensue even in cases other than when the mentioned term is being displayed in all of its glory, with those barely noticeable punctuation signs on either side of it. The well-known story of ‘niggardly’ has generated a lively political debate on either side of the spectrum.<sup>6</sup> Regardless of the position one may wish to take in that debate, its existence indicates that there is an issue to be discussed – that is, that even accidental or only remotely related tokens may need to be handled with care. I have heard that ‘donkey’ became a popular substitute for ‘ass’ due to the desire to avoid expressions phonetically close to ‘arse’ (Hughes 1991, 19). I do not know whether that story is true, but it is not in principle unintelligible.

I think it was Kaplan who once remarked that decent semanticists ought to be able to spell out the meaning of xenophobic slurs in non-xenophobic language. Indeed they can: ‘Kraut’ designates Germans and it expresses

---

<sup>5</sup> On the idea of semantic inertia see, for instance, (Davidson 1979) and (Cappelen and Lepore 2007); for my own views on pure quotation, see (Predelli 2009).

<sup>6</sup> See the *New York Times*, 31 January 1999. Incidentally, in Bolinger’s contrastive analysis of slurring and offensiveness, “to use the quotation name [e.g. “fuck”] rather than an available alternative [e.g. ‘the f-word’] ... warrants offense” (Bolinger 2017, 452).

disdain towards them, I may say. The details may be incorrect, but my attitude is untarnished, since I did not indulge in the use of the *analysandum*, as I would in standard Tarskian biconditionals such as: ‘snow is white’ is true iff snow is white. Similarly, I can surely provide a terse and not at all impolite analysis for ‘fuck’, as in, say: ‘fuck’ designates sexual intercourse and belongs to the coarse register. And yet, ‘fuck’ and its ilk put the student of language in a perilous situation. The referent, sexual intercourse, and the coarseness with which it is designated, are apparently neutralized when that word is displayed within quotation marks. Still, that very display suffices for the presence of those additional effects.

Something then remains that, metaphorically speaking, scopes out even of pure quotation. I refer to this aspect of an expression’s profile as its *taboo* status, and I proceed with a few remarks on taboo in the remainder of this essay.

### 3. Taboo tokens

Our normative assessments of utterances generally involve a worldly dimension, side by side with the genuinely linguistic features of an expression. So, ‘London is in France’ may be chastised as being false because it says that London is in France, because London is not in France, and because the speaker was in a situation where true talk was required. ‘Io ti ho invitato a pranzo’ may be criticized as impolite when directed to an adult stranger because, by virtue of its Italian meaning, ‘ti’ is familiar, and because, in our society, adult strangers are not to be addressed in a familiar register. Or else, *mutatis mutandis*, ‘Angela is a Kraut’ elicits our disapproval because it derogates the Germans and because, in reality, derogation is unjustified or downright impermissible. And so, at least in principle, you may try to escape those accusations while maintaining your allegiance as an English or Italian speaker by attempting to change the world: you relocate London, you protest that you were not really asserting anything about actual geography, you promote a reform of our attitudes towards adult strangers, or you argue that anti-German prejudice is justified.

In some sense, taboo is a more direct affair: the criticism incurred by the violation of a taboo simply ensues from the fact that, in the relevant

linguistic community, that word is being designated as taboo. You may well rejoice in that criticism, and you may enjoy your status as a rebel. But, if you utter a taboo word, you do not cease to be a target of disapprobation unless you relinquish your position as a member of that linguistic community.

That is not to say that the taboo status of an expression is utterly disentangled from the truth-conditional and/or non-truth-conditional aspects of its meaning. Presumably, some communities decree that, say, certain words for the divinity are taboo because they are privileged expressions for a being whose metaphysical status far exceeds that of anything that is a proper subject for human conversation. That is, they identify taboo expressions for reasons eventually having to do with the nature of their *designatum*, that is, for reasons related to their truth-conditional meaning. Or else, perhaps, ‘fuck’ ended up as taboo partly because of its connections with the non-truth-conditional aspects of its meaning, in a seamless shift from expressions that ought not to be used in polite company to words that should not even be mentioned. Indeed, in all likelihood, racial slurs such as the n-word derive their forbidden status both from the hideous historical oppression suffered by its designatum, and from the venomous history of derogation in the Western world. In this respect, it is hardly an accident that ‘Kraut’, ‘wop’, or ‘limey’ did not follow its destiny, and that they remained derogatory but non-taboo expressions.

And so, taboo may well occasionally (or perhaps even inevitably) have ensued from historical processes grounded on an expression’s *designatum*, on its non-truth-conditional meaning, or on something of both sorts. Yet this much is not equivalent to the denial of a distinctive niche for the conventional dimension of taboo. The contrast with truth-conditional meaning is patently obvious: truth-conditionally indistinguishable expressions may differ in their taboo status, as in the case of unobjectionable uses of ‘copulate’ for sexual intercourse. And so, certain noteworthy properties of the object or action or event that is being spoken about may perhaps have played a role in the generation of taboo, but they hardly suffice. The matter is no less clear when it comes to non-truth-conditional meaning: ‘screw’, in its use for sexual intercourse, is (or may well be) as coarse as ‘fuck’, but it is not (or need not be) taboo. And ‘spook’ is (or may well be) as derogatory

as the n-word, but its mention hardly elicits the same sort of sentiments. And so, coarseness, derogation, or other potentially alarming facets of non-truth-conditional meaning may lie behind the history of certain terms as taboo, but they do not bear the entirety of the explanatory burden.

More importantly, the conceptually distinctive features of taboo remain in place even if the contrasts mentioned in the foregoing paragraph are called into question on empirical grounds. And so, taboo's *reasons* are varied, and are in no way limited to expressions that deal with delicate subjects, or that begin their life in the coarse or derogatory arenas. They may, in fact, be utterly arbitrary motives, letting one expression go while forbidding the occurrence of one that is *fully* synonymous with the first, that is, while prohibiting a word that is indistinguishable from the truth-conditional and non-truth-conditional level alike.

The arbitrariness of taboo extends to our very *interaction* with the incriminated words. Certain practices forbid the spoken occurrence of an expression while allowing its being written down, perhaps as in Maimonides' intimation that 'only in the Temple is the name [of God] recited as it is written' (Laws of Prayer and Priestly Blessings, 14:10). I cannot claim competence with the proper interpretation of the great Sephardic philosopher, but it is at least conceivable that what he was after is the intimation that a certain word may harmlessly be tokened in its written form, but that it may not be spoken, or at least not 'recited as it is written'. Even more surprisingly, the relationships between an expression and its involvement in taboo may have to do not with the prohibition that it be pronounced, written, or tokened in any other form, but rather that it be *erased*, as in a proscription presumably stemming from no lesser source than *Deuteronomy*: 'you shall destroy [Pagan gods'] names from this place, [but] do not do this to God' (Deuteronomy 12, 3–4).

These scenarios may well be rather distant from the sort of linguistic dictates with which most of us are familiar. Yet, they are instructive to a much greater extent than the giggles engendered by the repeated mention of 'fuck' among a group of adolescents. What is instructive, in particular, is the deliberate sloppiness of my exposition in the paragraph above. Surely, if what taboo forbids is, say, the *spoken* occurrence of an expression, it is not that expression itself that is that practice's main target, but its verbal



tokens. Equally surely, expressions are not even the sort of thing that can be erased, since what is erased is a token, not the expression itself. And so, the taboo status of an expression must be interestingly related to its tokening, in ways that are more direct and illuminating than the limited significance of tokens in the study of its semantic meaning.

Returning to more familiar instances, then, the Kaplan-inspired project of a neutral analysis of coarse or derogatory terms inevitably breaks down when it comes to taboo. Or, at the very least, it breaks down when that analysis is being carried out: my speaking of the properties of ‘fuck’ need not indulge in an undesirable register, but it violates the dictates of taboo. And it does so inevitably, since any pronouncement as to the properties of ‘fuck’ must start with a specification of its object, that is, with the production of the incriminated token. Similarly, taboo’s apparent ‘scoping out’ of quotation results from the unavoidable layout for quotational terms. So, “xrt” (the result of appending quotation marks to ‘xrt’) refers to a three-letter string precisely by virtue of the fact that ‘xrt’ occurs within it, that is, by virtue of the fact that it is tokened as part of the tokening of that five-character affair. And “fuck” mentions a certain English four-letter word precisely by virtue of displaying that four-letter sequence as its proper part. Quotation may well be a fruitful tool for unveiling certain characteristics of taboo, but what does the trick, there, is the very format for its realization, the occurrence of the incriminated form.

It is thus unsurprising that the accidental occurrences that I have mentioned above, as in the case of ‘niggardly’, may cause alarm to those sensitive to the taboo status of certain expressions. The ailurophobic may have no qualms with ‘catatonic’, but those who attribute taboo status to ‘cat’ may legitimately opt for a different designation of unresponsive stupor. And, in relation to the aforementioned distinctions between different forms of tokening, one’s distress may be generated by (accidental or not so accidental) occurrences that are only phonetically in the vicinity of the incriminated term. I have heard of an overly sensitive teacher of French who refrained from revealing to her students the noun for seals, ‘phoque’, due to its phonetic vicinity with ‘fuck’. The sacrifice of her pedagogical mission on the altar of a minor taboo is unforgivable, but her motive remains comprehensible.

#### 4. Taboo and semantics

In the first parts of this essay I alluded to the distinctive role of taboo as a part of an expression's conventional profile that is independent from what I called its 'semantic meaning', that is, a part of meaning unamenable to the application the standard tools of semantic analysis. The idea that taboo is intimately connected with the process of tokening helps to shed some light on this vague idea.

In a famous passage in *Demonstratives* Kaplan highlighted the distinction between the occurrence of an expression (that is, the "combination of an expression and a context") and an utterance of it: the idea of an utterance comes "from the theory of speech acts," whereas the notion of a sentence-in-a-context derives "from semantics" (Kaplan 1989a, 522). Kaplan's use of 'speech acts' is closely related to the act of speaking, that is, to the event of tokening. Indeed, the reason why that distinction is important is that the semantic study of meaning ought to abstract away from the accidental regularities governing tokens. For instance, "utterances take time and utterances of distinct sentences cannot be simultaneous," but this physical inevitability ought to be kept at bay for the purposes of semantics: "we do not want arguments involving indexicals to become valid simply because there is no possible context in which all the premises are uttered, and thus no possible context in which all are uttered truthfully" (Kaplan 1989a, 522). This idea is reiterated in his later commentary, *Afterthoughts*, in terms of a telling mentioned slogan: "semantics [is] concerned not with the vagaries of actions, but with the verities of meaning" (Kaplan 1989b, 584–5). Once again, the features he deems to be *persona non grata* in semantic society derive from the physical or metaphysical structure of tokening: "utterances take time, and are produced one at a time; this will not do for the analysis of validity" (Kaplan 1989b, 584).<sup>7</sup>

And yet, there is no reason why features of this sort, though apt examples of Kaplan's methodological recommendation, ought to exhaust the domain for 'the theory of speech acts'. And so, by all means, let us ground

---

<sup>7</sup> For my developments of Kaplan's methodological advice, see (Predelli 2005) and (Predelli 2013).

our study of truth-conditional and non-truth-conditional meaning on occurrences, rather than utterances. And let us do so in order to abstract away from, *among other things*, the facts that utterances take time, that they involve the exercise of the mouth or the hand, or that they are inevitably performed by intelligent beings. But more besides may need to be swept aside: these non-conventional, natural features of the tokening act need not be all there is that needs to be expunged from the proper domain for semantics. In particular, other aspects of the tokening process may well ensue from conventional injunctions, in particular from decisions that do not affect tokening *in general*, but the tokens of *particular expressions*. Taboo, a conventional aspect of certain expression, apparently falls on that side of the divide, and it firmly belongs among “the vagaries of action” rather than to “the verities of meaning.”

A clear counterpart of all of this are the different targets for our normative assessment of an utterance. Those that are likely to catch the semanticist’s attention are those ensuing from the properties associated with occurrences, that is, eventually, with sentences in a context. From the perspective of *Demonstratives*, those are the properties ensuing from *character*, that is, from truth-conditional meaning. But there are no reasons why non-truth-conditional features ought not to feature here as well. And so, as mentioned at the beginning of this essay, an utterance may be chastised for being false, redundant, or contradictory. Or else, for being derogatory, in the wrong register, or impolite. And if it is censured for any of these reasons, it is so because it is an utterance representable in terms of a sentence that is primarily responsible for bearing those meaning-related properties. The discussion of the relationships between taboo and the tokening process indicates that matters are different when it comes to this phenomenon. There may well be nothing objectionable in calling the taboo status of an expression a part of its conventional profile, at least in the sense of being an aspect of that needs to be mastered by its competent users. But there are reasons for resisting a fully-fledged commitment to ‘meaning’ in this case: what is at issue is a convention that pertains not to an expression *in abstracto*, but to the action of tokening it.

In this sense, the study of taboo falls squarely within a normative theory of action, albeit, in our case, the sort of action that is of interest from the

viewpoint of language. Or else: a particular aspect of the conventional profile of certain expressions has little to do with anything that is of semantic concern, be it propositional encoding, truthful description, or the sort of expressive outcomes apparently engendered by virtue of non-truth-conditional meaning. This conventional dimension is rather fully captured from the viewpoint of a theory of linguistic action, that is, a theory of the sort of inter-personal coordination characteristic of conversational exchanges.

### 5. Concluding remarks: the power of taboo

The classic framework for the study of conversational moves focuses on their effects on common belief, that is, on a certain class of propositions. For instance, an assertoric and literal utterance of *s* apparently results in the proposal that *p* be added to common belief, where *p* is the proposition encoded in *s*.<sup>8</sup> The literature on non-truth-conditional meaning has contributed to an expansion of this model, generally directed towards a more nuanced and structured picture of conversation. For instance, an utterance of ‘Angela is a Kraut’ may well engender effects related to the speaker’s disdain for Germans, but it arguably proposes that these effects be recorded at a level other than that appropriate for an utterance of, say, ‘Germans are intrinsically unworthy of respect’.<sup>9</sup>

The details here are important and independently interesting, but they may safely be set aside here. What matters in the picture sketched above is an assumption that affects the truth-conditional and non-truth-conditional aspects of the uttered expression: the act of tokening intervenes in the economy of conversation precisely insofar as it is an act involving the presentation of a meaning-bearing affair, in the semantic sense of that term. And so, a certain enrichment of the conversational record ensues from my token of ‘London is in England’, precisely because the truth-conditional meaning of what I uttered eventually yields a particular proposition. Or

---

<sup>8</sup> See (Stalnaker 1999) and (Stalnaker 2014), and the considerable related literature; on the idea of assertions as proposals, see (Farkas and Bruce 2010).

<sup>9</sup> For discussions of the multi-layered nature of conversation see, for instance, (Roberts 1996), (Portner 2007), (Farkas and Bruce 2010), and (Anderbois et al. 2015).

else, something of a different nature happens when I token ‘Angela is a Kraut’, but that too stems from my words’ semantic profile, in this case, at least in part, their non-truth-conditional meaning.

In this picture of conversation, then, the tokening process intervenes at best as something that is inevitable for creatures of our kind, but that is, in and of itself, of little significance. In other words: the manifest act of tokening is there all right, but it is there only qua manifest exemplification of an affair endowed with truth-conditional and/or non-truth-conditional meaning. The case of taboo, on the other hand, forcefully invites us to reconsider our traditionally dismissive attitude towards the nitty-gritty of the marketplace: the sheer token may well be unaccompanied by any noteworthy semantic effects, as when it is merely mentioned, and yet reverberates with all of its force.

A thing done cannot be undone. From this metaphysical triviality comes the potency of taboo. Recall, as a term of contrast, the classic take on assertoric conversational moves as proposals of propositional enrichment. They are proposals that, clearly, may not in the end make it to common belief: ‘that is false’, you protest, thereby preventing our exchange from taking that claim on board. Non-truth-conditional affairs are notoriously harder to resist: as a rejoinder to ‘Angela is a Kraut’, your ‘that is false’ merely questions my attribution of nationality to that woman. And yet, thankfully, the xenophobe is not all-powerful: ‘hey, wait a minute, that is not the way to characterize the Germans’, you may protest.<sup>10</sup> And so, in either case, all can be undone, either by appealing to standard conversational tools such as denial and rejection, or by questioning the suitability of certain conversational developments.

Not so with taboo, whose power far exceeds that of non-truth-conditional meaning. ‘Fuck’, you utter. And, from the viewpoint of taboo, there is no taking it back, since the source of taboo, namely the very act of tokening, is what it is and cannot be undone. The prevention of taboo calls for the most literal form of silencing, that is, for forcibly preventing that token to occur in the first place.

---

<sup>10</sup> For applications of ‘hey wait a minute’ to issues of presupposition, see (von Fintel 2004).

All of the above surely deserves fuller theoretical attention. I wrap things up with a modest conclusion, independent of many details in my vague gestures thus far. For me, taboo and its conversational role turn out to be a profitable object of inquiry, both for their independent interest and for their repercussions on those other matters of semantic interest, most notably coarseness and derogation. More importantly, taboo also exerts an interesting pressure on our customary understanding of the study of language. Even if its role in human society turns out to be of lesser urgency than many other areas of semantic inquiry, the metasemantic repercussions of taboo deserve to be studied with care.

### Acknowledgements

Many thanks to the participants in the symposium Value in Language, hosted by the Slovak Academy of Sciences (March 29–31, 2021). A particularly emphatic thank you goes to Rob Stainton, who provided extensive and detailed comments on a first draft of this paper.

### References

- Allan, Keith. 1990. "Some English Terms of Insult Invoking Sex Organs: Evidence of a Pragmatic Driver for Semantics." In *Meanings and Prototypes. Studies in Linguistic Categorization*, edited by S. L. Tsohatzidis, 159–94. London: Routledge.
- Allan, Keith, and Kate Burridge. 2006. *Forbidden Words: Taboo and the Censoring of Language*. Cambridge: Cambridge University Press.
- Amaral, Patricia, Craig Roberts, and E. Allyn Smith. 2007. "Review of *The Logic of Conventional Implicatures* by Chris Potts." *Linguistics and Philosophy* 30: 707–749. doi.org/10.1007/s10988-008-9025-2
- Anderbois, Scott, Adrian Brasoveanu, and Robert Henderson. 2015. "At-issue Proposals and Appositive Impositions in Discourse." *Journal of Semantics* 32 (1): 93–138. doi.org/10.1093/jos/fft014
- Anderson, Luvell, and Ernie Lepore. 2013. "Slurring Words." *Noûs* 47 (1): 25–48. doi.org/10.1111/j.1468-0068.2010.00820.x
- Bolinger, Renée Jorgensen. 2017. "The Pragmatics of Slurs." *Noûs* 51 (3): 439–62. doi.org/10.1111/nous.12090
- Cappelen, Herman, and Ernest Lepore. 2007. *Language Turned on Itself. The Semantics and Pragmatics of Metalinguistic Discourse*. Oxford: Oxford University Press.

- Cruse, D. A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Davidson, Donald. 1979. "Quotation." *Theory and Decision* 11: 27–40. Reprinted in *Inquiries into Truth and Interpretation*, 79–92. Oxford: Clarendon Press.
- Davis, Hayley. 1989. "What Makes Bad Language Bad?" *Language and Communication* 9 (1): 1-9. [doi.org/10.1016/0271-5309\(89\)90002-5](https://doi.org/10.1016/0271-5309(89)90002-5)
- Diaz-Legaspe, Justina, Chang Liu, and Robert J. Stainton. 2020. "Slurs and Register: A Case Study in Meaning Pluralism." *Mind and Language* 35 (2): 156–82. [doi.org/10.1111/mila.12236](https://doi.org/10.1111/mila.12236)
- Dummett, Michael. 1973. *Frege: Philosophy of Language*. London: Duckworth.
- Farkas, Donka F., and Kim B. Bruce. 2010. "On Reacting to Assertions and Polar Questions." *Journal of Semantics* 27 (1): 81-118. [doi.org/10.1093/jos/ffp010](https://doi.org/10.1093/jos/ffp010)
- von Fintel, Kai. 2004. "Would You Believe It? The King of France is Back! (Pre-suppositions and Truth-Value Intuitions)." In *Descriptions and Beyond*, edited by Marga Reimer, Anne Bezuidenhout, 315–41. Oxford: Oxford University Press.
- Geurts, Bart. 2007. "Really Fucking Brilliant." *Theoretical Linguistics* 33 (2): 209–14. [doi.org/10.1515/TL.2007.013](https://doi.org/10.1515/TL.2007.013)
- Gutzmann, Daniel. 2015. *Use-Conditional Meaning*. Oxford: Oxford University Press.
- Harada, S. I. 1976. "Honorifics." In *Syntax and Semantics Vol. 5, Japanese Generative Grammar*, edited by Masayoshi Shibatani, 499–561. New York: Academic Press.
- Harris, R. 1987. "Mentioning the Unmentionable." *International Journal of Moral and Social Studies* 2: 175–88.
- Holmes, Janet. 1992. *An Introduction to Sociolinguistics*. London: Longman.
- Hom, Christopher. 2008. "The Semantics of Racial Epithets." *The Journal of Philosophy* 105 (8): 416–40. [doi.org/10.5840/jphil2008105834](https://doi.org/10.5840/jphil2008105834)
- Hughes, Geoffrey. 1991. *Swearing: A Social History of Foul Language, Oaths, and Profanity in English*. Oxford: Blackwell.
- Kaplan, David. 1989a. "Demonstratives." In *Themes from Kaplan*, edited by Joseph Almog, John Perry, and Howard Wettstein, 481–563. Oxford: Oxford University Press.
- Kaplan, David. 1989b. "Afterthoughts." In *Themes from Kaplan*, edited by Joseph Almog, John Perry, and Howard Wettstein, 565–614. Oxford: Oxford University Press.
- Kaplan, David. 1999. "What is Meaning? Explorations in the Theory of Meaning as Use." (ms)
- Kratzer, Angelika. 1999. "Beyond Ouch and Oops: How Descriptive Content and Expressive Meaning Interact." (ms)

- McCready, Elin. 2010. "Varieties of Conventional Implicature." *Semantics and Pragmatics* 3: 1–57. [dx.doi.org/10.3765/sp.3.8](https://doi.org/10.3765/sp.3.8)
- Portner, Paul. 2007. "Instructions for Interpretation as Separate Performatives." In *On Information Structure, Meaning, and Form. Generalizations Across Languages*, edited by Kerstin Schwabe, and Susanne Winkler, 407–25. John Benjamins.
- Potts, Christopher. 2007. "The Expressive Dimension." *Theoretical Linguistics* 33 (2): 165–98. [doi.org/10.1515/TL.2007.011](https://doi.org/10.1515/TL.2007.011)
- Predelli, Stefano. 2005. *Contexts: Meaning, Truth, and the Use of Language*. Oxford: Oxford University Press.
- Predelli, Stefano. 2009. "The Demonstrative Theory of Quotation." *Linguistics and Philosophy* 31: 555–72. [doi.org/10.1007/s10988-008-9042-1](https://doi.org/10.1007/s10988-008-9042-1)
- Predelli, Stefano. 2013. *Meaning without Truth*. Oxford: Oxford University Press.
- Richard, Mark. 2008. *When Truth Gives Out*. Oxford: Oxford University Press.
- Roberts, Craige. 1996. "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics." In *OSU Working Papers in Linguistics No. 49. Papers in Semantics*, edited by Jae-Hak Yoon, and Andreas Kathol, 91–136. The Ohio State University.
- Sauerland, Uli. 2007. "Beyond Unpluggability." *Theoretical Linguistics* 33 (2): 231–6. [doi.org/10.1515/TL.2007.016](https://doi.org/10.1515/TL.2007.016)
- Simons, Mandy, Judith Tonhauser, David Beaver, and Craige Roberts. 2010. "What Projects and Why." In *Proceedings of SALT 20*, edited by Nan Li, and David Lutz, 309–27.
- Stalnaker, Robert. 1999. *Context and Content*. Oxford: Oxford University Press.
- Stalnaker, Robert. 2014. *Context*. Oxford: Oxford University Press.
- Williamson, Timothy. 2009. "Reference, Inference, and the Semantics of Pejoratives." In *The Philosophy of David Kaplan*, edited by Joseph Almog, and Paolo Leonardi, 137–58. Oxford: Oxford University Press.
- Zwicky, Arnold. 2003. "The Other F Word." *Out* 115: 82–4.
- Zwicky, Arnold M., Peter H. Salus, Robert I. Binnick, and Anthony L. Vanek. 1971. *Studies Out in Left Field. Defamatory Essays Presented to James D. McCawley on the Occasion of His 33rd Or 34th Birthday*. Edmonton: Linguistic Research Inc.