Slurs in speech and thought

Tristan Thommen

To cite this version:

Tristan Thommen. Slurs in speech and thought. Psychology. Université Paris sciences et lettres, 2018. English. NNT : 2018PSLEE013 . tel-01887200

HAL Id: tel-01887200
https://theses.hal.science/tel-01887200

Submitted on 3 Oct 2018

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'École Normale Supérieure, Paris

Slurs in Speech and Thought

Les Péjoratifs dans le Langage et la Pensée

Ecole doctorale n°158

CERVEAU, COGNITION, COMPORTEMENT

Spécialité SCIENCES COGNITIVES

COMPOSITION DU JURY :

Mme. BIANCHI Claudia
Université Vita Salute San Raffaele
Rapporteur et présidente du jury

M. MARQUES Teresa
Université de Barcelone
Rapporteur et membre du jury

M. MUREZ Michael
Université de Nantes
Membre du jury

M. RECANATI François
École Normale Supérieure
Directeur et membre du jury

M. SPECTOR Benjamin
École Normale Supérieure
Directeur et membre du jury

Soutenue par Tristan THOMMEN
le 19 juin 2018

Dirigée par **François RECANATI**
et **Benjamin SPECTOR**

TABLE OF CONTENTS

ACKNOWLEDGEMENTS 8

TABLE OF FIGURES 10

WARNING 11

CHAPTER 1. INTRODUCTION 13

1.1. INTRODUCING SLURRING TERMS 14

1.2. GENERAL OUTLINE 25

1.3. EXPLANANDA 29

1.3.1. CENTRAL EXPLANANDA 29

1.3.2. PERIPHERAL EXPLANANDA 35

1.4. PROJECTION 37

1.5. THE HETEROGENEITY OF HYBRIDITY 46

CHAPTER 2. PRESUPPOSITIONAL ACCOUNTS OF SLURRING TERMS 54

2.1. "PRESUPPOSITIONALISM" 56

2.2. INTRODUCTION TO A PRESUPPOSITIONAL ACCOUNT OF SLURRING TERMS. 64

2.3. PROJECTION UNDER SIMPLE FILTERS 70

2.3.1. DISJUNCTIVE FILTERS 70

2.3.2. SUBJUNCTIVE CONDITIONAL FILTERS 74

2.4. PROJECTION UNDER COMPLEX FILTERS 79

2.4.1. DISJUNCTIVE FILTERS WITHOUT IGNORANCE IMPLICATURES 79

2.4.2. PARAPHRASES 81

2.4.3. NEGATED CONJUNCTIONS 83

2.5. CONCLUDING REMARKS 86

CHAPTER 3. TWO OTHER HYBRID ACCOUNTS OF SLURRING TERMS 88

3.1. CONVENTIONAL IMPLICATURE ACCOUNTS 89

3.2. SPEECH-ACT/CONVERSATIONAL IMPLICATURE ACCOUNTS 96

3.2.1. SPEECH-ACT ACCOUNTS OF SLURRING TERMS 96

3.2.2. NUNBERG'S VIEW 98

3.3. TOWARD A NEO-GRICEAN NUNBERGIAN ACCOUNT OF SLURRING TERMS 102

3.3.1. CANCELLABILITY 102

3.3.2. A NUNBERG-LIKE ACCOUNT 105

3.4. THREE CLARIFICATIONS 108

3.4.1. PRIMARY INTENTIONS AND SECONDARY INTENTIONS 108

3.4.2. MEANING FOR USERS, MEANING FOR OTHERS 109

3.4.3. NATURAL MEANING, AFFILIATION AND DE-AFFILIATION 111

3.4.4. TOWARD A THREE-DIMENSIONAL NUNBERGIAN ACCOUNT OF SLURRING TERMS 114

3.5. SLURRING TERMS WITHOUT ALTERNATIVES 116

3.6. THREE LIMITATIONS OF THE NEO-NUNBERGIAN ACCOUNT 119

3.6.1. COCOONED COMMUNITIES 119

3.6.2. EXPRESSIVITY 121

3.6.3. AUTONYMIC CONNOTATION AND CONVENTIONALIZATION 123

CHAPTER 4. TOWARD AN ACCOUNT OF SLURRING CONCEPTS 128

4.1. FROM COMMUNICATION TO THOUGHT 129

4.2. PRELIMINARY METHODOLOGICAL REMARKS AND CLARIFICATIONS 137

4.3. UPDATED EXPLANANDA 140

4.3.1. UPDATED CENTRAL EXPLANANDA 140

Cognitive role 142

Connections with STs 143

Central and Parasitic Cases 144

4.3.2. UPDATED PERIPHERAL EXPLANANDA 145

4.4. UNDERSTANDING THE CENTRAL/PARASITIC DISTINCTION 147

4.5. NORMAL CONDITIONS 149

CHAPTER 5. CALLING HYBRIDITY INTO QUESTION: INTRODUCING T-TERMS... 151

5.1. S-TERMS AND HYBRIDITY 153

5.1.1. FREGE'S VIEW ON HYBRIDITY 153

5.1.2. DISENTANGLING SUBJECTIVITY AND COMMUNICABILITY 157

5.2. "CUR" AND "JALOPY" AS T-TERMS 161

5.3. A REFERENCE-BASED VIEW OF EVALUATION IN T-TERMS AND CONCEPTS 164

5.4. THICK CONCEPTS AND NONSEPARATIONISM 169

5.5. OBJECTIONABLE THICK CONCEPTS VS. SEPARATIONISM 174

5.5.1. CONTEXTUAL FLEXIBILITY 174
5.5.2. INTERSUBJECTIVE FLEXIBILITY 176
5.5.3. BACK TO SEPARATIONISM 178
5.6. TWO UNSUCCESSFUL OBJECTIONS TO A REFERENCE-BASED VIEW OF EVALUATION........180
5.6.1. BACK TO PROJECTION 180
Objection 180
Reply to the Objection 182
5.6.2. CO-EXTENSIONAL CONCEPTS WITH DIFFERENT VALENCES 183
Objection 183
Reply to the Objection 185
5.7. SLURRING CONCEPTS AREN'T T-CONCEPTS 186
5.7.1. AMBIGUITY 186
5.7.2. REDUNDANCY AND CONTRADICTION 189
CHAPTER 6. A RADICAL NON-HYBRID ACCOUNT OF SLURRING CONCEPTS........ 193
6.1. INTRODUCING TCSE AND TCE AND TCA 194
6.2. THE PROBLEM OF PROJECTION 200
6.3. METALINGUISTIC NEGATION 203
6.3.1. "NON-F" 205
6.3.2. "IT'S FALSE THAT" 207
6.3.3. "WITHOUT BEING" 208
6.4. INTERNAL, EXTERNAL, AND METALINGUISTIC 210
6.5. ABSURD COUNTERFACTUAL CONDITIONALS 213
6.6. DEROGATION AND OFFENSE 215
6.7. CONCLUDING REMARKS 219
CHAPTER 7. A RESPONSE-DEPENDENT ACCOUNT 223
7.1. FROM ERROR THEORIES TO REALISM 225
7.1.1. COLOR SCIENCE AND ERROR THEORIES 226
7.1.2. ALTERNATIVES TO ERROR THEORIES 228
7.1.3. THE DISPOSITIONAL APPROACH 233
7.1.4. THE METAPHYSICS OF RESPONSE-DEPENDENCE 236
7.1.5. RD PROPERTIES AND RD CONCEPTS 239
7.1.6. RESPONSE-DEPENDENCE AND RELATIVISM 241
7.2. INTRODUCING RESPONSE-DEPENDENT ACCOUNTS 244

7.2.1. DISPOSITIONAL EFFECTS 244

7.2.2. DISPOSITIONAL CONCEPTS 245

7.2.3. GLOBAL RESPONSE-DEPENDENCE 249

7.3. SLURRING CONCEPTS MODELED ON RED 251

7.3.1. OPACITY 252

7.3.2. REFLEXIVITY........ 255

7.3.3. APPLICATION OF THE RED MODEL........ 257

7.3.4. REFLEXIVITY AND INDEXICALITY 263

CHAPTER 8. PROS AND CONS OF THE RED MODEL 266

8.1. EXPLANATORY PROS OF RDARED8........ 267

8.2. EXPLANATORY CONS OF RDARED8........ 278

8.2.1. THE PROBLEMS OF EXTENSION AND PHENOMENOLOGY 278

8.2.2. THE PROBLEM OF CIRCULARITY........ 279

8.2.3. CONCEPTUAL RESPONSE........ 281

1) Possession conditions........ 281

2) Fineness of grain........ 283

3) Categorization 284

8.3. APPEALING TO STEREOTYPES? 288

8.3.1. THE SURFACE STEREOTYPE VIEW (SSV) 289

8.3.2. PROS OF THE STEREOTYPE VIEW 291

8.3.3. CONS OF THE STEREOTYPE VIEW 292

8.4 FROM RDARED TO RDAPOLITE 294

CHAPTER 9. A NON-REFLEXIVE RESPONSE-DEPENDENT ACCOUNT........ 296

9.1. REFLEXIVITY AND NORMALITY 297

9.2. NON-REFLEXIVE CONCEPT POSSESSION........ 300

9.2.1. RECOGNITIONAL CONCEPTS........ 300

9.2.2. NON-REFLEXIVITY........ 305

9.3. SLURRING CONCEPTS MODELED ON POLITE 310

9.3.1. INTRODUCING RDAPOLITE........ 310

1) Control and coordination........ 312

2) A prioricity........ 312

3) Non-responder possession 313

9.3.2. HOW RDAPOLITE DEALS WITH THE CRITICISMS TO RDARED 314

9.4. REMAINING LIMITS OF RDA 316

CHAPTER 10. AN ESSENTIALIST ACCOUNT 323

10.1. ESSENCES 324

10.2. PUTNAM'S VIEW OF ESSENCES 327

10.2.1. OBJECTION 1: UNNECESSARY PROPERTIES 328

10.2.2. OBJECTION 2: UNNECESSARY ESSENCE 330

10.3. ARISTOTLE'S VIEW OF ESSENCES 334

10.4. THE ESSENTIALIST ACCOUNT OF SLURRING CONCEPTS AND THE EXPLANANDA 338

GENERAL CONCLUSION 344

APPENDIX 347

APPENDIX TO 3.3: THREE REMARKS ON NUNBERG'S ACCOUNT 348

CIRCULARITY? 348

IDENTIFYING THE RELEVANT GROUP 349

EXPRESSIVITY DOES NOT REDUCE TO THE RECOGNITION OF AN INTENTION 351

APPENDIX TO CHAPTER 5: THICK TERMS VS. SLURRING TERMS 352

APPENDIX TO CHAPTER 7: PERSPECTIVAL EFFECTS 358

A COLLECTION OF SLURRING TERMS AND OTHER INSULTS 362

GLOSSARY 365

REFERENCES 372

Acknowledgements

First and foremost, I would like to express my profound gratitude to my advisors François Recanati and Benjamin Spector for their invaluable input and precious guidance.

Special thanks to Michael Murez for the infallible support and devoted attention he gave me throughout these years.

I am very grateful to Claudia Bianchi and Teresa Marques. I am honored that these two exceptional scholars have accepted to be part of my committee.

I would also like to thank the following scholars, who dedicated some time and thought to my work. Many thanks to Aidan Gray, Anne Reboul, Brent Strickland, Esa Diaz-Leon, Carlo Geraci, Chris Barker, Claudia Bianchi, Emmanuel Chemla, Frédérique de Vignemont, Jack Spencer, Jeffrey King, Isidora Stojanovic, Judith Tonhauser, Geoffrey Nunberg, Giuliano Torrengo, Mark Richard, Max Kölbel, Michael Murez, Paul Egré, Philippe Lusson, Philippe Schlenker, Pierre Jacob, Peter Gärdenfors, Robert May, Robin Jeshion, Salvador Mascarenhas, Uriah Kriegel, Valeria Giardino, Vincent Homer, and Yael Sharvit.

Many thanks to the audience and commentators in the conferences, workshops and seminars in Gargnano, Paris, Milan, Barcelona, Uconn for their interest, and their interesting and sometimes challenging input.

I would also like to thank Alessandro Dupré, Alex De Carvalho, Alexander Martin, Alice Castan, Amandine Phok, Amir Anvari, Andreas Heise, Andrès Soria-Ruiz, Anna Giustina, Anouch Bourmayan, Antoine Collomb-Clerc, Armando Lavalle Terron, Arnaud Lazard, Arthur Huiban, Arthur Portron, Aurélien Nioche, Aurélien Schwartz, Aurélien Tonneau, Auréliane Pajani, Axel Barrault, Benjamin Icard, Bianca Cepollaro, Blaise Farrokhi, Brice Bantegnie, Camille Rozier, Camille Thorey, Charlotte Vandendriessche, Christina Ioannou, Christophe Gardella, Claire Pelofi, Clara Arnoux, Clémence Almeras, Clémence Vazard, Clémentine Fourrier, Dalia Cohen, Dan Hoek, David Landais, Deniz Ozyildiz, Diego Feinmann, Emma Vilarem, Emmanuel Noblins, Eric Tromer, Félix Larrieu, Filippo Contesi, François Le Corre, François Monin, Gabriel Sulem, Géraldine Carranante, Gerda Ana

Melnik, Guillaume Dezecache, Guillaume Dumas, Helena Hachmann, Hélène Lestage, Henri Vandendriessche, Hernan Anllo, Hippolyte Gros, Hugo Mell, Ignacio Rebollo, Indrek Reiland, Inès Crespo, Insa Lawler, Jean Thorey, Jean-Louis Lorsold, Jean-Rémy Martin, Jennifer Lawlor, Jeremy Kuhn, Jérémy Zehr, Jérémie Lafraire, Joulia Smortchkova, Juliane Farthouat, Juliette Vazard, Juliette Perdreau, Lise Hobeika, Lorna Le Stanc, Lou Safra, Louis Rouillé, Luca Gasparri, Lucie Hobeika, Manuel Kriz, Mariana Rebelo, Marie Guillot, Marie Malafosse, Marine Guillanton, Marjolaine, Markus Kneer, Martin Fortier, Martha Vogel, Maryam Ebrahimi Dinani, Mathieu Kauffmann, Mathieu Koroma, Mélanie Lochkareff, Michael Kurrels, Milica Denic, Mora Maldonado, Nariaki Nakayama, Natalia Karczewska, Nathalie Evin Abitbol, Nathalie Marcinek, Neil Zeghidour, Nicolas Francillon, Nicolas Porot, Nicolas Vienney, Nura Sidarus, Palle Leth, Pierre Lazar, Raphaelle Malassis, Regina Subow, Reinaldo Bernal, Romain Bourdoncle, Rosalynd Koch, Samira Boujidi, Samuel Recht, Sandra Lasry, Sharon Peperkamp, Sophie Bilardello, Slawa Loev, Sofia Lismont, Sonia Marin, Sophie Berlioz, Sylvain Charon, Théodora Digidipoulos, Thibaud Griessinger, Thomas Pirot, Timothée Behra, Tricia Magalotti, Tomoo Ueda, Valentin Brunck, Valeria Mongelli, Victor Carranza, Victor Pitron, Victoria Aubert, and Vincent Gaudefroy.

Merci à mes amis Akim, Amandine, Aurélien, Florent, Gabrielle, et Jonas.

Une pensée pour mes parents Hélène et Patrick, pour ma soeur Ariane, et pour Abel.

Table of Figures

FIGURE 1. HOM AND MAY'S DEROGATION/OFFENSE DISTINCTION 216

FIGURE 2. FOUR CLASSES OF EVALUATIVE TERMS AND CONCEPTS 221

FIGURE 3. NINE VIEWS ON SLURRING TERMS AND CONCEPTS 221

FIGURE 4. A PHOTOGRAPH BY GEORGES ROUSSE 238

FIGURE 5. FROM THE REALIZER TO THE DISPOSITIONAL EFFECT........ 245

FIGURE 6. EXAMPLE OF FRAGILITY 245

FIGURE 7. SITUATING RESPONSE-DEPENDENT CONCEPTS AMONG DISPOSITIONAL CONCEPTS........ 247

FIGURE 8. A "NORMAL" SUSTAINING MECHANISM FOR RED 308

FIGURE 9. A DEFERENTIAL SUSTAINING MECHANISM FOR RED 308

FIGURE 10. FROM PROPERTIES TO COMMUNICATION........ 309

Warning

The present work investigates the nature of expressive meaning and focuses on slurring terms, hence I shall advise the reader that she or he will, throughout the manuscript, run into many occurrences of very derogatory and offensive words. Some authors prefer avoiding words targeting the most oppressed and disenfranchised groups, others avoid mentioning slurring terms altogether. I have preferred to be more direct and explicit.

Although I understand and am sympathetic to the discomfort some feel in writing down such words, I want to make clear that all slurring terms figure in what follows as bits of linguistic data, which I look at and theorize about from the outside perspective of the theoretician. Slurring terms are simply taken as linguistic objects whose effects are to be accounted for. I find it counterproductive and unnecessary to suppress these effects in artificially picking defused slurring terms, or replacing them with symbols, because it impacts the very phenomenon to be explained and makes it less patent.

Just like in any other scientific investigation, what is better than looking straight at a phenomenon to try and understand it? I prefer not to be shy about our object of study, and see no harm in doing so in the context of linguistic and philosophical work. I hope the reader will share my stance on the issue and will not be offended by the many occurrences of words loaded with hatred and bigotry.

Chapter 1. Introduction

Brief outline

In this introductory chapter, I introduce the reader to slurs and some of their central features. I introduce and give informal definitions of what I will call *Slurring Terms* (STs) and their so-called *Neutral Counterparts* (NCs), I distinguish the slur's extension from its "*Target*", I introduce the *Co-extensionality Thesis* and the family of *Hybrid expressivist* accounts. This sets the landscape for what will follow.

I then propose a general outline of the dissertation to give the reader a bird's eye view of the work. Next, I establish a list of all the central and less important features of slurring terms, in order to have a set of explananda with regard to which we can evaluate the different theories we will investigate throughout the manuscript. I give to the feature of projection a special status, because it is arguably the main linguistic property of these terms, a property they seem to share with many expressions and constructions in natural language.

I close the chapter with a series of remarks on the notion of hybridity, which seems central in the case of slurring expressions which appear to have hybrid content. I insist that other expressions and constructions in natural language display such hybridity of content, and propose a threefold distinction between kinds of hybrid content depending on their projection profile: Projective-layering, projective-filtering, and projective-expressive content. I thus pave the way for a first family of hybrid linguistic accounts of slurring expressions that I will focus on successively in what follows: presuppositional accounts, conventional implicature accounts, and accounts in terms of conversational implicatures.

1.1. Introducing Slurring Terms

On June 22nd 2015, a harsh polemic followed Barack Obama's mention of the n-word. I am referring to his words in an interview for the podcast WTF with Marc Maron:

> Racism, we are not cured of it. And it's not just a matter of it not being polite to say "nigger" in public.

It was clear to everyone who participated in the debates that the n-word was mentioned under direct quotation, not really used. Moreover, it was clear to everyone that Obama's comment was meant to be a contribution to anti-racism. But despite the consensus on quotation and on the truthfulness of Obama's anti-racist intentions, his mention of the n-word has been accused of causing offense, harming the targeted communities, perpetuating racism, and so on and so forth.

This fact is striking, for how could any of these acts have been performed by a mere mention of a word *under quotation marks*? How can anything leak out of pure quotation? If a judge wishes to condemn someone for her repeated use of the n-word for instance, can't she report the words of the defendant without thereby causing harm and offense to the targets?

Some data seem to suggest, on the contrary, that it is not necessarily derogatory to mention words like the n-word. Here is a small collection of citations displaying such cases where the force of slurs seems to be somehow defused:

> The fact is that people routinely produce sentences in which the attitudes implicit in a slur are attributed to someone other than the speaker. The playwright Harvey Fierstein produced a crisp example on MSNBC, "Everybody loves to hate a homo." Here are some others: "In fact We lived, in that time, in a world of enemies, of course… but beyond enemies there were the Micks, and the spics, and the wops, and the fuzzy-wuzzies. A whole world of people not us…" (edwardsfrostings.com)

> So white people were given their own bathrooms, their own water fountains. You didn't have to ride on public conveyances with niggers anymore. These uncivilized jungle bunnies, darkies... You had your own cemetery. The niggers will have theirs

> over there, and everything will be just fine. (Ron Daniels in *Race and Resistance: African Americans in the 21st Century*)
>
> All Alabama governors do enjoy to troll fags and lesbians as both white and black Alabamians agree that homos piss off the almighty God. (Encyclopedia Dramatica)
>
> [Marcus Bachmann] also called for more funding of cancer and Alzheimer's research, probably cuz all those homos get all the money now for all that AIDS research. (Maxdad.com)

In these examples, the speaker is using slurs, but at the same time seems to disapprove of the attitudes usually associated with the slurs she uses. This is a phenomenon that Bolinger coins "insulation":

> Consider a hypothetical corporate memo, advising employees that they must abide by a strict anti-slurring policy:
>
> MEMO: The following terms are not to be used by any Corp. employee, nor is their use to be tolerated in any Corp. classroom or workspace: 'chink', 'dyke', 'honky', 'nigger', 'spic'… [etc.]
>
> It is doubtful that anyone would protest that the slurs as they occur in the memo are as offensive as they would be if they were simply used. So to at least some extent, mentioning slurring terms successfully insulates their offense potential. (Bolinger 2015, p. 5)

But Bolinger adds:

> There is still something strange (or offensive) about listing each of the slurs explicitly rather than giving a blanket admonition to avoid slurring terms. If so, that suggests that something other than a simple use/mention distinction is at work in mitigating, though not entirely neutralizing, the offensive potential of these terms. (Bolinger 2015, p. 5)

This feature of words like the n-word - their apparent ability to leak out of pure quotation - is so striking and unusual that it, in itself, deserves careful attention. We will see throughout the present work other features that motivate a theoretical interest in these pejorative representations.

Terms like the n-word form a lexical category that has not been much explored by linguists and philosophers of language in the past, despite the foundational questions that some of their most central features seem to raise. The field is now becoming aware of such a shortcoming, so that an increasing number of researchers are turning their attention to them. They seem to belong to a class of terms that are used to convey an affective judgment, or an evaluative attitude, about the members of the category they denote.

Let us now try to informally isolate some of the fundamental distinctive characteristics of these devices in order to appreciate the nature of the difficulties they give rise to, and to eventually explore the conceptual grounds on which to build an adequate analysis of these features, with the linguistic and philosophical tools at our disposal.

Slurs are derogatory terms targeting groups or individuals on the basis of their ethnicity (e. g. "nigger"), gender (e. g. "cunt"), sexual orientation (e. g. "faggot") and the like. Puzzles about slurs arise from contrasts like the following[1]:

(1) !There is a Boche downstairs.

(2) There is a German downstairs.

An utterance of (1) will usually convey a derogatory content that an utterance of (2) will usually lack, and since (1) and (2) minimally differ in that "German" in (2) replaced "Boche" in (1), the predicates "Boche" and "German" must be distinct in content.

That two terms are distinct in content is of course not puzzling *per se*; it becomes puzzling when we observe that, even though "Boche" and "German" differ in content, they stand in a privileged relation to each other. It intuitively seems that the relation between "Boche" and "German" is tighter the relation between "Boche" and "Chinese", between "chink" and "nigger" or between "French" and "German" for instance.

We could suppose that this tighter relation is simply one of entailment. But if "Boche" asymmetrically entailed "German", then "Boche" would apply only to a subclass of

[1] I introduce the symbol "!" to mark the expressivity/offensiveness of utterances featuring slurs.

[2] Jeshion notices that "gook" has no available neutral counterpart. This term was used by the

Germans, and this does not seem to be the case. Germanophobes rather apply "boche" to all Germans. The nature of the relation between "boche" and "german" does not seem to be one of entailment then.

Slurs are of interest because - at least pre-theoretically - they seem to fulfill two roles at once: picking out a referent (or at least being *about* certain individuals), and expressing, or signaling, certain attitudes, or implicit beliefs, or emotions on the part of the speaker.

It is indeed a unanimously recognized feature of slurs that they have the power to express strong negative attitudes or affective judgments towards the members of a category. The sorts of attitudes that can be expressed by a slur come in a large variety, just as the intensity of their offense varies across uses (e. g. "nigger" is said to be more offensive than "chink").

We could try to identify these attitudes with contempt, derogation, dismissiveness, hostility, hate, disgust, fear and so on; such a precise description of complex emotions should be the object of experimental work, I won't enter here in all the details of these intricate feelings and attitudes that slurs express and convey.

But if the mere expression of a negative attitude were sufficient to capture the phenomenon constituted by slurs, "nigger" and "kike" wouldn't be that different, because they are both used to express negative attitudes.

Slurs in fact also target individuals and groups of individuals *on the basis of* their ethnicity, religion, gender, political or sexual orientation, appearance, life style, profession and so on. Individuals are targeted by the strong negative attitudes expressed by slurs *because* they belong to a certain category.

For example, the slur "frog" can be used to display, let's say, *contempt* for French people, and to do so on the basis or their membership in the class of French people. Hence, slurs seem to have a classificatory function in addition to their function to express attitudes, and that's why they have been linked to *thick terms* (like "chaste" or "lust") whose alleged specificity is precisely to mix classification and attitude.

This dual function of slurs is observable in their coextensivity with a possible "neutral" term whose use refers to the same individuals or groups of individuals without offending or being derogatory (The slur "frog" thus has "French person" as its *neutral counterpart*).

> Every slur, so far as I can tell, has or could have a "neutral counterpart" which co-classifies but is free of the slur's evaluative dimension (Richard, 2008: 28)

Even when there is no other term available in the language to refer to, or at least to be about, the same group of individuals than the slur[2], it is nevertheless always possible to conceive of a term that would retain the classificatory function of a slur and drop the attitude conveyed with its use.

Moreover, this dual property of slurs is highlighted by the apparent synonymy between what I will call *analytical* dissociated slurring expressions, like "dirty Jew", and *integrated* slurring expressions, like the term "kike".

Analytical slurring expressions clearly dissociate a classificatory function, located in my example in the neutral term "Jew", and an attitudinal part, here located in the modifier "dirty". This analysis can be easily extended to intensified expressions like "fucking Jew" or "damn Jew", whose difference with slurs also seems to be syntactic.

> It is as if someone who used, say, the word "nigger" had made a particular gesture while uttering the word's neutral counterpart. An aspect of the word's meaning is to be thought of as if it were communicated by means of this (posited) gesture. (Hornsby 2001, p. 11)

In what follows, primarily to avoid confusion between slurring speech acts and the terms that are used to perform them, I will call slurs "slurring terms" (STs). I will call the terms STs respectively stand in the relevant privileged relation to their "neutral counterparts" (NCs):

> **Slurring Terms (STs)**[3]: terms whose meaning is seemingly *hybrid* (it is made of at least two different kinds of meaning) and whose meaning components are *separable* (one can find or construct neutral counterparts)

[2] Jeshion notices that "gook" has no available neutral counterpart. This term was used by the United states military during the Vietnam and Korean wars for their east-Asians enemies (Jeshion, 2013a),

[3] All technical notions I introduce and discuss are summed up in the glossary p. 365.

STs thus seem to be *hybrid*, in the sense that they have meanings of two different kinds. In addition, the two types of meaning must be *separable*, as one can observe with the existence of neutral counterparts to STs, sharing one but not the other type of meaning:

> **Neutral Counterparts (NCs)**: A representation is a Neutral Counterpart (NC) of a hybrid representation when it shares its descriptive component and lacks its attitudinal component. For instance, "jew" is the NC of the ST "kike".

I will later critically examine the alleged separability of STs. That "boche" and "German" stand in a privileged relation to each other is corroborated by the five following observations.

(i) STs differ from NCs.

To be convinced of (i), simply compare (1) and (2). An important *explanandum* will thus be to account for the nature of the relation between a slurring term and its neutral counterpart (if there is any).

(ii) STs have content.

We should exclude a logical possibility which would explain in which sense STs differ in content with their neutral counterpart: the possibility that STs don't have content.

The notion of content is far from being clear and precise, but it should be noted that STs clearly have use conditions. This can bee seen with utterances like (3), where it is clear that something wrong must have happened for it to be performed:

(3) !Vladimir Putin is a boche.

An utterance of (3) either is a Germanophobic way to make the false statement that Vladimir Putin is German, or else a sign that the speaker is simply ignorant about an important aspect of the word "boche".

The fact that "boche" bears content is unanimously recognized, but there is wide disagreement with regard to what kind(s) of content exactly it carries, how it is represented in the speaker's and hearer's mind, and conveyed in communication.

(iii) STs carry projective derogatory content.

Pairs like the following illustrate this seemingly additional power that STs, and certainly many other expressions in natural language, possess. Although (4a) and (4b) might at a first glance seem to be nearly synonymous, they display crucially distinct behaviors under their negated alternatives, respectively (5a) and (5b).

(4) a. !John is a faggot.

b. John is homosexual and worthy of contempt because of that.

(5) a. !John is not a faggot.

b. John is not homosexual and/nor worthy of contempt because of that.

Whatever offensive and derogatory content the ST "faggot" conveys, it still conveys it under negation in (5a), whereas (5b) is a neutral and non-problematic statement.

This is unexpected from the point of view of a theorist who would try to identify the derogatory effects of STs with the ascription of a complex derogatory predicate to a subject. Indeed, negating that an individual has a certain complex property should not be particularly derogatory. (5a) is then wrongly predicted to be innocuous.

The same phenomenon can be observed under operators other than negation, such as conditionals (6), modals (7), questions (8), quantification over events (9) etc.

(6) !If Mary met a kike, so did her father.

(7) !Mary must have met a kike.

(8) !Did you meet a kike?

(9) !Every time Mary meets a kike, her father is sad.

Such data is usually taken to show that there is something in the content of STs that is not affected by truth-conditional operators, and to set the goal of incorporating this additional *projective* derogatory content in our theories of language. I will later discuss the notion of projection and how it applies (or not) to STs.

(iv) Not all STs are synonymous.

There are different STs in the language: "nigger" targets black people, "spade" targets black people too (it was mostly used in the sixties in hippy communities), "faggot" targets homosexuals, and so on and so forth. The relation between different STs having the same target, like "nigger" and "spade", is controversial, but all agree that "nigger" and "faggot" are not synonymous.

In other terms, the difference between "nigger" and "faggot" is way more dramatic than the difference between "nigger" and "spade". This is an important thing to remark, because an account of the expressive powers of STs will thus have to leave a place for STs to differ from one another, even though they are similar in that they express negative attitudes or emotions.

> (v) STs have targets.

We saw that the words "nigger" and "faggot" are both derogatory STs and still differ in content. Intuitively, "nigger" and "faggot" differ with regard to the set of people these expressions are *about*, or are directed at. Whether or not it should be called a "reference", STs have a "target", that is, they somehow pick out a group of individuals in the world[4]. As we shall see later on, there are reasons to deny that STs refer altogether, such as the intuitive falsity of slurring statements.

> **Target**: The target of a slurring representation is the group or individual it is meant to apply to.

Note that Neutral Counterparts (NCs) could alternatively be defined as sharing their target with STs rather than their "descriptive component". This gives us a new definition of NCs:

> **Neutral Counterparts' (NC's)**: A representation is a Neutral Counterpart (NC) of a hybrid representation when it shares its target and lacks its attitudinal component.

In what follows, I will talk about "NCs" interchangeably, as the difference between the two definitions will not crucially matter.

(i)-(v) taken together suggest that *STs target different groups and carry distinctive projective derogatory contents*. It therefore seems natural to argue that the relevant relation between STs and their NCs is a relation of co-extensionality. This is what I call the co-extensionality thesis:

[4] See Richard (2008), Hom (2008) and Hom and May (2013, 2015), for an alternative.

> **Co-Extensionality Thesis (CET)**: STs have the same extension as their NCs.

Note that CET is to be understood not as the thesis that STs and NCs have the same extension as a matter of luck or coincidence in the actual world, but that they have the same extension *necessarily*, in every possible world (the same *intension*). CET is indeed meant to be a view on the *meaning* of STs, so that a competent user who believes that "John is a Boche" must also believe that "John is German", and the other way around - provided the user in question is Germanophobic. The nature of the co-extensionality is thus like that of "oculist" and "ophtalmologist", and not like that of "rational animal" and "bipedal animal that does not have feathers".

Facing CET, the simple and natural theory of STs goes as follows. If a slurring term does target the same class of people as its neutral counterpart but still differs from it (e. g. in that it carries an additional power to offend or to display attitudes and emotions), then there must be additional dimensions of content at play. This natural theory can thus be said to be *hybrid*.

An expressive dimension on top of a descriptive dimension in slurring terms would account for the difference in content between the ST and its neutral counterpart. Hybrid views of slurring terms therefore let the truth-conditions of STs do the reference-fixing job, and then call on other dimensions of meaning or other theoretical tools to account for their additional expressive properties:

> **Hybrid Expressivist Accounts (HEA)**: Hybrid expressivist accounts of STs subscribe to the CET and call on other dimensions of meaning to account for their additional expressive properties.

Since slurring terms would do a dual job, we ought to have a dual story of their functioning.

Although the CET thesis is controversial, there is even more debate with regard to the handling of the so-called *expressive dimension*. Different sorts of potentially useful distinctions between two dimensions of meaning have been drawn in the course of decades of research on the human language faculty, and we could in principle count as many hybrid attempts at modeling the alleged hybridity of STs.

Frege introduced distinctions between *sense* and *tone* (1979, posthumous writings) to deal with contrasts such as the one between "dog" and "cur", or between *sense* and *reference* (1879) to deal with contrasts such as the one between "Hesperus" and "Phosphorus". Austin

(1975) introduced a distinction between propositional *content* and *force* to deal with speech-acts of questioning, ordering, asserting etc. Kaplan (1977) introduced a distinction between *character* and *content* to deal with indexicals and demonstratives. Grice (1975) introduced a distinction between *what is said* and *what is implicated* to clarify the semantic/pragmatic divide. Strawson (1950) inspired by Frege (1892), introduced a distinction between what is *at issue* and what is *presupposed* to distinguish truth-conditions from failure-conditions.

All these distinctions, among others, between two dimensions of meaning in natural language could in principle be used to develop a candidate hybrid model of slurring terms and begin confronting it to empirical data. We will investigate several versions of HEA in the next two chapters.

Before going any further, note that there are a number of related terms whose analysis would shed light on STs: expressive intensifiers like "damn" or "fucking", interjections like "shit!" or "crap!", laudatives like "sweetheart" or "saint", thick terms like "lewd" and "courageous" and so on. One general task of the analyst is to understand the syntactic, semantic, and pragmatic differences between these classes of terms. It is thus worth situating an investigation of STs in a broader landscape of expressives and pejorative expressions[5].

A particularity of STs, as opposed to personal insults like "jerk", is that they target individuals in virtue of their belonging to a certain group (like the Germanophobic insult "boche" for instance), whereas personal insults - or "particularistic pejoratives" - target individuals in virtue of their own personal characteristics. Saka notes that

> the difference between particularistic pejoratives [e. g. "jerk"] and slurs does not lie in their denotative functions but in the affective attitude of the speaker toward an individual versus toward a class of individuals. (Saka 2007, p. 149)

Among all the terms that seem at the same time to refer to a group or an individual and to express an attitude, there is no *a priori* reason to treat STs as *sui generis*. All expressions with the relevant features could in principle be gathered under a broader notion of "S-terms"

[5] The reader will find a collection of slurring terms and related expressions towards the end of this dissertation.

(STs), but I will most often focus on the particular case of slurs in the discussions about "STs". Ideally, everything I will be led to say shall be applicable to the broader class of STs.

Before developing and evaluating different possible accounts of STs, it is important to identify more precisely a full set of distinctive features and characteristics they have, and that should be accounted for. Before turning to that, I will present a general outline of the dissertation. I will then dedicate the remainder of this chapter to specifying the major explananda in light of which we will evaluate the various accounts that I will put forward and investigate, and to discuss a bit further the notion of hybridity so as to be able to move to a first version of a hybrid view in the following chapter: a presuppositional account of slurring terms.

1.2. General Outline

The present dissertation is aimed at better understanding slurs, their structure, their function(s), their cognitive underpinnings, and the theoretical lessons we could draw from their existence in natural language. Intuitively, slurs are pejorative terms targeting groups or individuals, and they have a seemingly hybrid content - with a descriptive and an attitudinal component.

A first task I will handle, in the remaining of this introductory chapter, is to come up with a set of explananda with regard to which we can evaluate the different possible accounts I will investigate. We will discuss and define major properties of slurs such as projection and expressivity, but also many other interesting features they have.

I will eventually discuss in more depth the notion of hybridity, and see many other expressions and constructions in natural languages displaying a similar sort of hybrid content. In particular, we will see that presuppositions, conventional implicatures, and conversational implicatures constitute promising bases for a hybrid linguistic account of slurring expressions. This is why I will successively explore each of the three views.

Chapter 2 is dedicated to the exploration of a presuppositional account of slurring expressions. I discuss the notion of presupposition, present what a presuppositional account of slurs should look like, and put forward data that such accounts cannot handle, thus falsifying the view. Presuppositional accounts of slurs predict too narrow a projective profile for slurs.

In chapter 3, I present and discuss the next two hybrid linguistic accounts of slurs: the conventional implicatures (CI) account and an account in terms of conversational implicatures. I will show that a CI account of slurs is well armed to derive all the linguistic data - except for a potential contrastive behavior in responses to questions -, but that a more general, theoretical objection can be raised against it. I come back to this general objection later, because it can in fact be addressed to all purely linguistic accounts of slurs.

I then investigate in depth a conversational account which I think is the best on the theoretical market, Nunberg's account in terms of manner implicatures and affiliatory speech-acts. I discuss Nunberg's account and reformulate it in a Gricean manner so as to account, potentially, for even more phenomena than the initial account. I show that the account is extremely robust, except for a blind spot having to do with the possibility of slurs without counterparts in an isolated community of slurrers. The investigation of these three main hybrid linguistic accounts of slurs will lead me to formulate a general objection to linguistic accounts of slurs.

In chapter 4, I develop my objection, which is based on the observation that hybrid accounts, even though they are descriptively adequate (as the conventional implicature account might be), lack a clear theoretical framework. Describing a linguistic phenomenon is one thing, explaining it is another. My aim is more to explain and understand than to predict and describe. I argue that an important dimension has been neglected in the debates surrounding slurs: psychology. Why do slurs have the properties they have? They project in such and such a way that we can describe in such and such a manner, but why?

I make the bet that slurs derive most of the interesting properties they have from features of a mental representation, a concept, that they are used to express. After a couple of necessary terminological and theoretical clarifications, I update our earlier set of explananda consequently, and dedicate the remaining of the dissertation to pursuing an account of what I call "slurring concepts".

Chapter 5 is an attempt at building a first view of slurring concepts by questioning one of my starting hypotheses: the hypothesis that slurring terms and concepts are *hybrid*, that is, that their semantics contains two dimensions. The point of the chapter is to investigate whether slurs conform to the hybrid S-term model or another model I put forward: the T-term model. The introduction of such terms helps me put forward a reference-based account of the evaluative content of slurs.

This view argues that T-terms and concepts are, appearances notwithstanding, not truly evaluative: they simply have a rich descriptive content such that they refer to subgroups, subgroups which are independently, extra-semantically evaluated as being negative. The evaluation ends up being associated with the terms, but it becomes associated only extra-semantically. The semantics of these terms is thus one-dimensional.

I attempt at connecting the debate to an existing literature on so-called "thick" terms and concepts, which raise similar questions having to do with two potentially separate dimensions of meaning. I then address a series of (unsuccessful and then successful) objections to the resulting view, enriching the descriptive content of slurring terms and concepts but keeping the evaluativity as an external element.

Chapter 6 then explores a more radical theory of slurring concepts locating all of their dimensions, including the evaluation itself, in the truth-conditional layer. According to such a view, that I call the "truth-conditional account", slurring concepts are simply complex descriptions such as "worthy of contempt because of...".

Based on joint work with Cepollaro, I discuss such a view and show that it will always have a hard time accounting for projection facts. Based on novel data, I show that the data usually put forward to deal with projection (e. g. "There are no kikes") are confounded by metalinguistic factors, I argue.

In the next three chapters (chapters 7, 8, and 9), I thus move to another approach to slurring concepts, which appeals to what some call "response-dependent" concepts. I aim, in several ways, at an account of slurring concepts as a species of response-dependent concepts. Response-dependent concepts - typically secondary quality concepts such as RED - have the interesting property of being inherently connected to non-conceptual, purely cognitive responses. Moreover, their extension is determined via the possessor's sensitivity to certain features of her environment. These essential properties of response-dependent concepts make them excellent candidates for a reduction of slurring concepts.

Thus, in chapter 7, I present the notion of response-dependence, develop two important notions of *opacity* and *reflexivity*, and develop a response-dependent account of slurring concepts based on the model of RED.

In chapter 8, I assess the pros and cons of this account with regard to its ability to handle our updated list of explananda. We will face the need to add some complexity in the response involved in slurring concepts so as to explain the apparent categorization behavior of possessors which crucially differs from that of possessors of RED. Indeed, possessors of RED rely on their perceptual response to categorize an object as red, whereas it is unlikely that racists similarly rely on their racist response to categorize their targets as members of the target group.

I will then explore the possibility of giving to the notion of stereotype a role in the response itself, but we will see that this is not fully satisfactory, because categorization does not seem to rely on statistics. This discussion - among other minor issues such as a potential circularity of our definition of slurring concepts -, will lead us to consider another response-dependent account which gives up the property of reflexivity.

After a discussion on reflexivity and non-reflexivity, I develop in chapter 9 such a non-reflexive response-dependent account, based on the model of concepts such as POLITE. According to this view, it is possible to possess a slurring concept even in the absence of the right sort of cognitive response. Although it addresses some of the problems raised against the first response-dependent account, we will see that such an account looses track of the initial reason we had to invoke response-dependence, which was the inherent link between the concept and the non-conceptual response.

I will then develop a potentially more satisfactory account of slurring concepts as essentialist concepts. Under successively two understandings of the notion of essence - one modal and another Aristotelian -, I will put forward the view that slurring concepts postulate an essence in their targets, and that this essence is taken to have a negative value. The combination of these two theses, gathered under the name "Essentialist account", has the resources to account for most, if not all, of the explananda we started with, I shall argue.

The reader will find in the appendix a (non-exhaustive) list of slurring terms and other related expressions, a glossary (p. 365) gathering the terminological notions and theories I introduce throughout the dissertation, and a few additional remarks on issues discussed in chapters 3, 5, and 7 (on Nunberg's account, on thick terms, and on perspectival effects).

1.3. Explananda

Here I present and discuss here all the major explananda for slurring terms. Note that I will eventually be led in to update this list of explananda, based on the need to postulate what I will call "slurring concepts" on top of slurring terms.

The major question that theorists are interested in evolves around the so-called "expressivity" of slurring terms, but there are many other desiderata that a theory of STs should meet. Some of the explananda below come from the literature on slurs, others are extra desiderata I think should be added.

Here are all the explananda I could think of for slurring terms, split in two categories depending on whether they are central explananda or peripheral explananda. The present work focuses and attempts at giving an adequate explanantia only of the central explananda.

1.3.1. Central Explananda

- *Expressivity*. In many important cases, STs seem to involve an emotional or affective content, or to feature an evaluative component that is hardly reducible to their descriptive semantic value. Any theory of STs should give an account of the nature of this intuitive "expressivity" of STs, and if possible, to connect them with the "expressivity" of other expressives in natural language that are not STs. As we saw, this might be the feature of STs (with projection, below) that was given the most attention so far.

- *Projection*. The fact that the expressivity of STs scopes out of most semantic operators is considered by many theorists to be one of the main properties of STs to account for. I thus dedicate the whole following section to this crucial explanandum.

- *Offense/Derogation*. Uses of STs offend and derogate targets and bystanders. This is a simple observable feature of STs that shall be accounted for. Not all terms offend and derogate, we thus need to understand how certain terms seem to have acquired this particular power.

- *Defectiveness*. I think that there is a strong sense in which STs are morally flawed terms. It is obvious to all but to some users of slurs that these terms are in some sense wrong, that using them is not morally insignificant, as shown by the strong moral reactions they tend to elicit. A theory of STs should strive to locate their main intuitive defect(s). Here, I pursue the working hypothesis that STs are not merely morally or ethically flawed, but that they also involve cognitive flaws, and correspond to an inappropriate way of categorizing and reasoning about reality.

Note that this explanandum has a very special status, because talk of "inappropriate" or "defective" ways of categorizing is normative, and in this sense does not have the scientific neutrality we expect from an investigation on a class of terms. I think it is worth giving pride of a place to it though, because as theorists, we all intuitively recognize that these terms are flawed, and clarifying in which sense exactly we take them to be flawed might be enlightening about their nature and specificities as linguistic entities.

- *Neutral Counterparts (and extension)*. We saw that STs were taken to have NCs. This comes with two related explananda. The first is that a proper account of STs should explain how they are alike, and how they differ from one another. Additionally, seemingly coextensive STs such as "kike" and "yid" must differ in their cognitive roles, and a proper account of STs should individuate them so as to allow for that fact. The second explanandum coming with the notion NCs is to say what the extension of STs is - leaving open the possibility that they have a null extension and simply share their "target" with NCs.

- *Dehumanization and identifying thinking*. Jeshion (2013) argues that one of the main role of STs is to encode dehumanizing modes of thought. The cognitive act of dehumanization should certainly be clarified. It is clearly involved in racism, as can be seen in the common theme likening people to animals (Jahoda, 2015, also see the list of slurs in the appendix). See Haslam (2006) for an integrative review on dehumanization.

Uses of STs also seem to classify the targets so as to reduce their identity, as if being e. g. a "boche" was what the target really *is* (again, see Jeshion, 2013). Note that the identifying component of STs seems linked to expressivity, because it seems that speakers succeed in identifying their targets through the expression of their derogatory attitudes, such as regarding the targets as fundamentally inferior.

- *Derogatory variation*. Some STs are more offensive than others (Hom, 2008). There is inter-group variation (for instance, the anti-Semitic "kike" is judged to be more insulting than the outdated Germanophobic "boche") and intra-group variation (for instance, the racist n-word is way more offensive and pejorative than the somewhat rare "spade", although the two terms have the same target). An account of offensiveness and expressivity should thus allow for such offensiveness and expressivity to come in degrees.

- *Ideologies and Stereotypes*. Uses of STs seem to convey, or to be associated with, racist ideologies and stereotypes about the targeted groups. There are at least four independent reasons to believe that stereotypes play a non-negligible role in shaping STs:

> *i)* STs simply tend to bring stereotypes to mind (Jeshion 2013b).
>
> *ii)* Stereotypes also dehumanize and harm the target's self-conception (Jeshion 2011).
>
> *iii)* Stereotypes and STs are both associated with a taboo (Anderson & Lepore 2013)[6].
>
> *iv)* Derogatory variation might indicate that different stereotypes are associated with different targets.

An account of STs should be able to characterize the nature of the connection between STs and ideological and stereotypical thinking about the targets.

- *Contempt*. STs are closely related to negative moral emotions, such as contempt. The nature of the link between uses of STs and the relevant negative moral emotions should be clarified. Do emotions intervene in categorization, and if so, how exactly?

- *Reluctance to evaluate*. Another property of STs is our reluctance to attribute truth (or falsity) to 3rd person descriptive statements where they are used. For example, competent -

[6] Since I will not come back to this point, note that I take it to be unlikely that the prohibition imposed on STs is sufficient to account for all their features - as Anderson & Lepore (2013) seem to be aiming at. The main reason is that there exist languages with certain phonological forms that can express either a STs or another non-slurring term. For instance, the Italian "finocchio" is either a ST targeting homosexuals, or the name of fennel. What is it that could be prohibited in the former case and not in the later?

and non-racist - speakers typically feel uncomfortable when asked to evaluate whether (10) is true or whether (11) is false:

(10) !Angola is mostly inhabited by niggers.

(11) !Angola is mostly inhabited by chinks.

This reluctance is striking because intuitively, being inhabited mostly by black persons is a property correctly attributed to the republic of Angola in the first statement (so that the sentence - or the thought - should be true), and the property of being inhabited mostly by Chinese people is incorrectly attributed to the republic of Angola in the second statement (so that the sentence - or the thought - should be false).

But our mitigated intuitions as competent speakers with regard to the truth and falsity of such racist statements is surely a symptom of a complex interaction between different components of our linguistic and social competences, and any account of STs has to offer an explanation for this striking piece of data.

This phenomenon is reminiscent of presuppositions, and has motivated different versions of presuppositional theories of expressives (and of STs in particular) that I discuss below.

- *Derogatory autonomy*. Another significant feature of STs is the apparent autonomy of their derogatory force from the beliefs, attitudes, or intentions of their users (Hom, 2008).

It is never an option for me to use a derogatory word in a non-slurring fashion, even though I am full of good intentions and attitudes (letting aside special cases of *non-weapon* uses that I evoke below). Even in the actual absence of offensive or aggressive intent, uses of these expressions do harm. Mere uses of such words - again, except non-weapon uses - are independently capable of being offensive, in virtue of what seems to be their conventional content alone.

For instance, the aggressiveness and offensiveness triggered by the use of a ST or insult such as "you bastard" cannot be cancelled by simply adding "no offense" or "I don't mean to be insulting".

As Bolinger (2015) remarks after Culpeper (2011),

> it is possible to be offensively rude/impolite without even being aware that you have done so. (Bolinger, 2015, p. 7).

This feature is patent in cases of accidental derogation, such as Potts' (2007) tale of a new school superintendent attempting to stand against racism by saying that "Niggers come in all colors", thereby outraging his audience.

Even more, as we saw, it seems to require important contextual prerequisites to even mention these terms, to cite them in direct reports. For example, at least some speakers judge it borderline to metalinguistically correct a racist's utterance by uttering (12):

(12) !?There are no "chinks" at the *École Normale Supérieure*, only Chinese people.

If confirmed, the alleged autonomy of the derogatory powers of STs from the attitudes of the speakers is an important feature, because it hinges on the conventionality of the phenomenon. And if anything, we want to know whether and to what extent the phenomenon is conventional, and the theoretical consequences we could draw from this fact.

- *Various perlocutionary powers*. Slurring is a speech-act. In addition to classification and expression of negative attitudes, STs are also used to *do* different sort of things, to cause a variety of harm to the targeted groups.

We can try to identify these action potentials that STs seem to carry with insulting, denigrating, humiliating, stereotyping, belittling, dehumanizing, assaulting, making propaganda, subordinating, affiliating oneself with a group and so on, but such a precise description of complex acts one can perform with the *things* these words are is beyond the scope of the present work. I consider all aspects of slurring as a speech act to be beyond the scope of my thesis.

Nevertheless, accounting for STs should at least give us a hint about the way in which they come to have these different perlocutionary powers. After all, these powers are also reasons why some speakers use STs.

- *Understanding*. It is likely that STs play a role in allowing most speakers, even those that do not use STs, to understand STs. Hom & May (2013) take this question to be the central question about STs:

> How can a competent, rational speaker of a language know the meaning of a pejorative without being committed to, or even complicit with, racist attitudes? (Hom & May 2013, p. 1)

It is interesting that every competent speaker of the language has the capacity to understand STs even without being disposed, in any situation, to use them. Indeed, everything looks as if STs were not part of the idiolect of non-racist speakers, because they are not terms they use, but they understand to some extent what the terms are about and perceive their offensiveness.

We shall thus distinguish between the meaning of STs as possessed by "normal" users, and the meaning of STs as possessed by the rest of us, because the two are not necessarily the same. This is linked to the following explananda.

- *Possession conditions*. A right account of STs should account for which subjects count as normal users in which conditions, in a sense of "normal" that should be independently clarified. This is important, because it is clear that there are different classes of users of these terms. There are the primary ones who possess and use STs because they are racists, and there are the rest of us, who somehow understand these terms but do not use them.

When we will be talking about concepts later on, the question of possession conditions will be even more important, because it will be the driving force behind projection and expressivity.

- *Central/Parasitic*. Uses of STs come in large variety, and not all of them are on a par. Some are central and others are incidental. There is indeed a variety of special uses of STs that we need to account for. I follow Jeshion's pervasive identification of such uses and the terminology she introduced (Jeshion, 2013a).

First, some uses of STs seem not to display any offensiveness (call these *non-weapon uses*), such as *appropriated* uses among targets. Appropriated uses are uses of STs by the members of the targeted group as a friendly way to call each other - "queer" and "nigger" have such uses nowadays.

Second, some uses of STs target a sub-class of the category denoted by their neutral counterparts, like in (13), and other uses target a larger class than the neutral counterpart, like in (14):

(13) !Obama is not a nigger. (*G-contracting* use)

(14) !Jefferson is a nigger. (*G-extending* use)

Standard uses of STs reference the same category of people as the neutral counterpart - call these *G-referencing* uses - and my discussion will be mostly focused on uses that are considered to be central: *G-referencing weapon uses* of STs, in order to better examine the problematic relations between their truth-conditional - i. e. descriptive - properties and their potential to perform acts of offense, insult, moral evaluation etc.

The task of pulling these two sorts of uses/deployments apart should be distributed among i) the necessary preliminary work of isolating the phenomena to be explained and ii) the explanation of the phenomena.

This explanandum indeed has a special status because its explanantia is somewhat theory-dependent, and it shall thus be treated separately. I will be doing so in Chapter 4.

1.3.2. Peripheral Explananda

Here are some explananda that I consider to be peripheral, but this is relative to what one is interested in of course.

- *Creation and evolution*. It appears to be very easy to create a ST contextually. Simply name the target group with the name of their favored food for instance, and you have a ST (e. g. "beaner" targeting Mexicans). Why and how do terms (so rapidly) come to be STs?

And why is their expressive power so sensitive to change over time, as shown by e. g. the amelioration of archaic STs like "boche" or "kraut"? Many STs were far more offensive in the past than they are today, and inversely, many STs which are extremely offensive today were arguably not that harsh just a handful of years ago. The meaning of non-offensive terms like "table" does not seem to change so rapidly, so why would STs be so sensitive to time?

- *Group formation, binding, and identity*. Uses of STs seem to play a non-negligible role in shaping communities. A trivial observation shows that sharing contempt for out-group members usually contribute to form and bind groups together, and STs could very well be used to play such a role.

- *Formation and perpetuation of social hierarchies, and of bigotry*. Jeshion (p. c.) reminds us that the expression of STs is also a device for building social hierarchies and for maintaining preexisting ones. They would also be useful to communities for the transmission of contempt toward the targets, hence for the perpetuation of bigotry.

- *Community endorsement*. Uses of at least some STs signal the community's endorsement of the bigotry expressed, which overtakes the speaker herself. It is not only the speaker's contempt that is expressed by uses of STs. In some cases, as we clearly see with the N-word, the harm and threat that a use represents exceeds whatever harm a single individual could do on her own. It is as if uses of the N-word signaled that the community endorses the bigotry and oppression that the speakers expresses. Such an observation is to the best of my knowledge due to Saka:

> Since the conventionalization of contempt relies, like all convention, on societally recognized norms, every pejorative utterance is proof not only of the speaker's contempt, but proof that such contempt prevails in society at large. This is why pejoratives make powerful insults, why repeated exposures to pejoratives can create feelings of alienation, inferiority, and self-hatred, and indeed why a single pejorative utterance evokes measurable bias in overhearers (Greenberg & Pyszczynski 1985, Kirkland et al. 1987, Simon & Greenberg 1996). (Saka, 2007, p. 142)

Note that on top of bringing about the endorsement of the community, it seems that uses of STs also seem to bring to mind histories of past oppression.

A noteworthy linguistic feature of STs is that, even when they are syntactically embedded under various operators, their expressivity and offensiveness has the ability to semantically "scope out": their use is offensive and insulting even when they are under the scope of negations (15), indirect reports (16), conditionalizations (17), modals (18), event quantifications (19) and so on.

(15) !Mary didn't meet a frog.

(16) !My father told me that Mary met a frog.

(17) !If Mary met a frog, then so did her father.

(18) !Mary must have met a frog.

(19) !Every time Mary meets a frog, her father is sad.

Potts coined this feature *nondisplaceability*, noticing that this particularity shows that expressives in general

> ...predicate something of the utterance situation. (Potts, 2007)

This particular feature of STs is linked to the difficulties we have to repudiate them directly, by denial. For example, we cannot correct a racist claim by merely uttering: "No! There are no niggers here" without being accused of racism ourselves.

What is it exactly for a content to be or not to be affected by truth-conditional operators? Are there different ways not to be affected by truth-conditional operators? Are there different possible causes for being insensitive to some operators?

The term "projection" was introduced to denote this fact - that some content carried by linguistic devices is insensitive to the compositional potential of standard operators and predicates, like negation, conditionals, modals, to name just a few. This feature of projection was first noticed in the case of presuppositions, to the extent that projection was at the time simply identified with presuppositional calculus. Any projective material was then analyzed as a presupposition.

It is now widely recognized that a whole range of diverse items or constructions display a projective behavior without necessarily being presuppositional, as we shall see below. This immunity to truth-conditional operators, which seems at first glance to constitute a clear failure of compositionality, and whose instances have been taken to argue for several versions of a multi-dimensional semantics (Potts, 2000, 2005, 2007, Nouwen 2006, McCready, 2010, Gutzmann, 2015)[7], seems to come in several varieties, triggered by a variety of linguistic patterns and devices.

But the reduction of projection to presupposition is indeed tempting, because projection seems to be closely linked to the notion of a *common background*, just like presuppositions. In rough terms, what projects is what must be common ground (Karttunen 1973, 1974) for the utterance to be felicitous. In some cases though, the projective content of a presupposition can be *accommodated* (Lewis, 1979), that is, added to the common ground in which it was absent prior the utterance.

A point to be stressed before trying to characterize and give a definition of projection, and then to study its presence in natural languages, is the following: in order to see that projection is a somewhat unexpected phenomenon, it is important to notice that the operators under which the triggers stand are normally used to cancel propositional content. These operators do not preserve the entailments of their arguments.

Focus on the so-called "family of sentences test": negating-(20b), conditionalizing-(20c), questioning-(20d) or modalizing-(20e) (20a) does cancel its propositional content. In every world in which (20a) is felicitous and true, Paul is happy; I note that "□P". Cancellation of (20a)'s propositional content therefore corresponds to "¬□P", that is, there exists at least one

[7] On the alleged non-truth-conditional meanings of projective content, see Bach, 1999 on conventional implicatures (of which he claims they cannot involve a degree of commitment from the speaker (speakers can be committed to CIs as much as to what is said), they cannot involve a degree of explicitness ("Yes", and "No" express implicit propositions, but are not CIs), and cannot be non-truth-conditional, because CIs are truth evaluable).

> In general, after all, utterances do not *communicate* [Bach's emphasis] that the conditions for their appropriate performance have been met. (footnote 8 p. 332)

accessible world in which the utterance is felicitous and true where P (that Paul is happy) is not the case.

(20) a. Paul is happy $\Box P$

b. Paul is not happy $\Box\neg P$ hence $\neg\Box P$

c. If Paul is happy, then so is Mary $\neg\Box\neg P$ and $\neg\Box P$, hence $\neg\Box P$

d. Is Paul happy? $\neg\Box\neg P$ and $\neg\Box P$, hence $\neg\Box P$

e. Perhaps Paul is happy $\neg\Box\neg P$ and $\neg\Box P$, hence $\neg\Box P$

With these three considerations at hand - i) logical entailments have a categorical behavior with regard to projection, ii) various entailments or inferences obtain in virtue of the tokening of triggers, and iii) operators of the "family of sentences test" are usually entailment-cancelling operators - we can attempt a first definition of projection.

Simons, Tonhauser, Beaver and Roberts's (2010) define projection in the following way, which I coin the *Immunity definition of projection* (IDP):

> **IDP**: "An implication projects iff it survives as an utterance implication when the expression that triggers the implication occurs under the syntactic scope of an entailment-cancelling operator."

IDP illustrates why projection is sometimes called "scopelessness".

Using the "family of sentences" test, we can start diagnosing projective behavior in several constructions of natural language. The goal is simply to consider lots of constructions under four standard environments (negation, if-clauses, questions, and modals) in order to let the data give us a hint on an appropriate taxonomy of projective content.

The ultimate goal of such a taxonomy would be to help us locate the projective behavior of STs in a wider landscape, to show that i) STs, even though they raise foundational questions on the very nature of meaning, do so on a par with plenty other expressions that it is worth considering in parallel, that ii) STs are nonetheless distinct from other species of projective triggers, and that iii) it is worth studying the functioning of STs since, in addition to teaching us about how meaning is encoded in language and conveyed through communication in virtue of their projective behavior, it will eventually allow us to develop a finer-grained analysis of the relations between speech and thought. I will come back to these issues throughout the present work.

Now, what are the kinds of things that project, and that we could attempt a reduction of STs to? Simons, Tonhauser, Beaver and Roberts (2010) proposed that the notion of projection was closely related to the notion of "not being the main point", i. e. so-called "not-at-issueness".

Karttunen and Peters (1979), as well as Horton and Hirst (1988), also described the behavior of *presuppositions* by noticing that they are propositions that a given utterance is not primarily about.

Similarly, Potts (2005) described *conventional implicatures* in noticing that they are "not the main point" of the utterance. What is it for a proposition that is somehow conveyed in a given context to "be the main point" of an utterance, or "not to be the main point" of an utterance? Do utterances convey contents in different layers and different nature, and how? Is there a way to diagnose whether a given proposition that is conveyed by an utterance in a context is part of the "main point"? Are there many different ways not to be the main point of an utterance, and if yes, what are the differences that can be observed?

Presuppositions seem to convey propositions that are "not the main point", to the extent that for a while, every proposition that seemed not to be "the main point" of an utterance tended to be identified with a presupposition. The above questions are thus important to understand the role that a theory of presupposition must have in a general theory of content.

To start with, here are six types of linguistic entities which pass the "family of sentences" test.

• Presuppositions

(21) a. Bob stopped eating meat.

b. Bob didn't stop eating meat.

c. If Bob stopped eating meat, he will eat more beans and peas.

d. Did Bob stop eating meat?

e. Perhaps Bob stopped eating meat.

Projects: Bob used to eat meat.

(22) a. My brother is happy.

b. My brother is not happy.

c. If my brother is happy, I'm happy.

d. Is my brother happy?

e. Perhaps my brother is happy.

Projects: The speaker has a brother.

(23) a. The queen of Germany is bald.

b. The queen of Germany is not bald.

c. If the queen of Germany is bald, she has a modern hairdresser.

d. Is the queen of Germany bald?

e. Perhaps the queen of Germany is bald.

Projects: Germany is a monarchy.

- Expressives (intensifiers, interjections, honorifics, register, conventional implicatures...)

(24) a. Alfred forgot his fucking keys.

b. Alfred didn't forget his fucking keys.

c. If Alfred forgot his fucking keys, he will be late.

d. Did Alfred forget his fucking keys?

e. Perhaps Alfred forgot his fucking keys.

Projects: whatever attitude is expressed by the use of "fucking"

(25) a. Professor Tarski wears a raincoat.

b. Professor Tarski wears a raincoat.

c. If Professor Tarski wears a raincoat, it must be snowing.

d. Does Professor Tarski wear a raincoat?

e. Perhaps Professor Tarski wears a raincoat.

Projects: respect towards Tarski

(26) a. Peter saw John's bellybutton.

b. Peter didn't see John's bellybutton.

c. If Peter saw John's bellybutton, he is a good physician.

d. Did Peter saw John's bellybutton?

e. Perhaps Peter saw John's bellybutton.

Projects: whatever childish connotations are associated with the term "bellybutton"

- Slurring terms (STs)

(27) a. !There is a kike downstairs.

b. !There is no kike downstairs.

c. !If there is a kike downstairs, we'd better leave.

d. !Is there a kike downstairs?

e. !Perhaps there is a kike downstairs.

Projects: Derogation towards Jewish people

• Supplements: Appositives (integrated non-restrictive relative clauses (NRRs with *qui* in french), non-integrated non-restrictive relative clauses (NRRs with *lequel* in french), appositive nominals (ANs)), parentheticals

(28) a. Napoleon (Suzie's cat) won the fight.

b. Napoleon (Suzie's cat) didn't win the fight.

c. If Napoleon (Suzie's cat) won the fight, he will have more offspring.

d. Did Napoleon (Suzie's cat) win the fight?

e. Perhaps Napoleon (Suzie's cat) won the fight.

Projects: Napoleon is Suzie's cat.

• Gestures (see e. g. Schlenker 2015)

(29) a. Luca punished$_{\text{[thumb-index pinch to the ear]}}$ his son.

b. Luca didn't punish$_{\text{[thumb-index pinch to the ear]}}$ his son.

c. If Luca punished$_{\text{[thumb-index pinch to the ear]}}$ his son, so did Mary.

d. Did Luca punish$_{\text{[thumb-index pinch to the ear]}}$ his son?

e. Perhaps Luca punished$_{\text{[thumb-index pinch to the ear]}}$ his son.

Projects: If Luca had punished his son, the punishment would have consisted in pinching his ear.

(30) a. Luca helped$_{\text{[both palms gently moving upward]}}$ his son.

b. Luca didn't help$_{\text{[both palms gently moving upward]}}$ his son.

c. If Luca helped his son, so did Valeria.

d. Did Luca help$_{\text{[both palms gently moving upward]}}$ his son?

e. Perhaps Luca helped$_{\text{[both palms gently moving upward]}}$ his son.

Projects: If Luca had helped his son, the help would have consisted in a form of push up.

• Effects of focus, stress, or intonation:

(31) a. Andreas gave A BOOK to Mary.

b. Andreas didn't give A BOOK to Mary.

c. If Andreas gave A BOOK to Mary, then she won't read it.

d. Did Andreas give A BOOK to Mary?

e. Perhaps Andreas gave A BOOK to Mary.

Projects: (roughly) Andreas gave something to Mary (the stress makes it not at issue that something or other was given, but there is a contrast between giving a book and giving something else - which are the open alternatives in the common ground).

(32) a. Andreas has given a book to MARY.

b. Andreas didn't give a book to MARY.

c. If Andreas gave a book to MARY, then she won't read it.

d. Did Andreas give a book to MARY?

e. Perhaps Andreas gave a book to MARY

Projects: (roughly) Andreas gave a book to someone.

(33) a. Andreas introduced Mary to SAM.

b. Andreas didn't introduce Mary to SAM.

c. If Andreas introduced Mary to SAM, everyone will be happy.

d. Did Andreas introduce Mary to SAM?

e. Perhaps Andreas introduced Mary to SAM

Projects: (roughly) Andreas introduced Mary to someone.

(34) a. Andreas introduced MARY to Sam.

b. Andreas didn't introduce MARY to Sam.

c. If Andreas introduced MARY to Sam, then everyone is happy.

d. Did Andreas introduce MARY to Sam?

e. Perhaps Andreas introduced MARY to Sam.

Projects: (roughly) Andreas introduced someone to Sam.

The above data show that there are at least six families of expressions and constructions that do display a projection behavior with regard to the family of sentence test: presuppositions, expressives, slurring terms, supplements, gestures, and effects of focus or intonation. But are there any differences in the projection behavior of these different devices, or can we merge some - or even all - of them under a single label?

There seems to be several ways not to be affected by negation and other truth-conditional operators, and several reasons for it. With the sole criterion of projection in the family of sentences test as defined above, we negatively gather in the same category a host of very diverse and complex phenomena: whatever the phenomenon consists in, if it is not sensitive to negation and other operators, then it *projects*, we are told.

But it is worth trying to clean up the muddle by introducing finer-grained distinctions among non-at-issue contents, so that we don't gather different patients under the same label and risk prescribing them the same treatment, though they suffer from different ills.

I propose here - with examples - a rough threefold-distinction among different kinds of projective inferences, depending on the reason why the contents in question are insensitive to truth-conditional operators: *i)* projective-filtering, *ii)* projective-layering, and *iii)* projective-expressive contents.

First, there are all the things that are insensitive to truth-conditional operators because they are not at all in the conventionalized content of the utterance. I call such projective content *projective-filtering* content.

Grice's natural meaning could be an instance of projective-filtering content. When someone performs an utterance with an Italian accent for instance, we can draw the inference that the speaker is Italian and this inference projects to the matrix. That is, if the word uttered with an Italian accent was embedded under any sort of truth-conditional operator, hearers are still entitled to make the same inference.

Second, there are all the things that are insensitive to truth-conditional operators because, even though they are encoded in the truth-conditions, they are presented as not being the main point of the utterance. I call these projective content *projective-layering* content.

An example is supplements. When a non-restrictive relative clause (NRR) specifies some information about the subject of an utterance for instance, like in "Napoleon, the French emperor...", the inference that Napoleon is the French emperor projects, but this is not because it is not conventionally encoded in the content of the expressions uttered. Quite the opposite. The information is linguistically encoded, it is just presented as if it was somehow whispered aside, not the main point, and is hence not captured by semantic operators operating on the at-issue level.

Third, there are all the things that are insensitive to truth-conditional operators because they are expressive/non-conceptual. I call it *projective-expressive* content.

An example is the expressive intensifier "fucking", which triggers the projective inference that, say, the speaker is in a heightened emotional state, but the information, although it is somehow linguistically encoded, does not seem to be truth-conditionally encoded. Projection in this case seems to come from the very nature of the encoded meaning, rather than from the way in which it is conveyed (as in the previous case).

Whether or not these categories could be merged, are empty, or should be spliced further should of course be discussed. The above threefold distinction is only a first working hypothesis addressing the need to clarify the notion of projective content. We will see that STs could in principle be located in either of these three categories of projective content, but that the resulting views are not equivalent.

Let us now take a closer look at the three above kinds of projective content. We saw that what projects is simply what is not affected by negation-like operators. Now, imagine that an inference is not part of the linguistic content of any of the uttered expression, that it is just present, independent of linguistic content.

For example, going back to the first of my three categories, the inference that my interlocutor is a human being, or the inference that she has well-functioning vocal organs, is independent of the meaning of the expressions uttered. A mere mumble or some piece of nonsensical voice-like sounds would bring up the same inferences. Of course, we expect such inferences not to be affected by negation-like operators.

In other words, these inferences might be unaffected by operators simply because they were never in the content of the expressions that were uttered in the first place. Such projective-layering inferences may be analyzed as pragmatic implications (Recanati, 2003), implications of an action, preconditions that must obtained for the utterance to be possible, and so on.

If a speaker uses the word "table" at any point in an utterance, we can draw the inference that, for instance, the speaker is not mute. This inference is therefore a piece of information that is conveyed, not by the conventional linguistic content of the utterance, but by the action of uttering itself.

Place the same word "table" under negation, and the utterance obtained still triggers the inference that the speaker is not mute. Such an inference therefore "projects" in the above sense: it passes the family of sentence test. But it is really different for an inference to project because it is a pragmatic implication, or a precondition that must be fulfilled for the utterance to be performed, than for it to project because of the special way it is conventionally encoded in the trigger. The expressive content of STs could project because of a similar sort of mechanism, as we shall see when I will discuss Nunberg's account in chapter 3[8].

Compare this case to the case of an utterance of "my sister" which, even under operators, carries the information that the speaker has a sister, or to an utterance of the expressive "damn" which, even under negation, conveys negative attitudes or emotions. Intuitively, such inferences really *depend* on content encoded in the word "my", or respectively "damn", as long as it is encoded in a way which is compatible with their specific projection profiles.

This first category of very broad projective content which are insensitive to operators because they are not in the relevant dimension of conventional, encoded content, I call *Projective-filtering* content. Under negation or other negation-like operators, such inferences still obtain, because these sorts of inferences rely not on our properly linguistic faculty, but either on our ability to read other people's minds or on very general reasoning mechanisms. Since such inferences are not the product of any linguistic mechanism, linguistic devices like negation do not affect them.

There are already several phenomena discussed that are good candidates for being part of this category. Kaplan (1999), Predelli (2013), or Recanati (1981), independently discuss different versions of a use-based semantics as a good alternative to truth-conditional

[8] Another difference seems to be about whether the inference is part of the communicative intention. Some inferences that are not encoded are part of the communicative intentions, and some have suggested this even for run-of-the-mill presuppositions.

semantics to deal with such phenomena in natural language, involving non-linguistic - and sometimes non-conventional - dimensions of content. Let us now consider some other cases of projective-filtering contents.

First, pragmatic implications - the implications of an action - project in this very broad sense I just identified. Just like Grice's (1957) natural meaning, pragmatic implications are preconditions that must be fulfilled for an action to be performed, so that an action being performed entails that its preconditions are met.

An utterance is indeed a kind of action, and thus has preconditions that must be met - such as not being mute, or armless in the case of signed languages. As we just saw, the inference that the speaker is not mute is one we can draw only after the speaker's utterance has begun. So it is an inference elicited by the utterance, but it is not part of the conventional linguistic content of the utterance.

A second example of what I call projective-filtering could be so-called "use-conditions". Consider the distinction there is in European French between two different second person singular pronouns: "tu" and "vous". Both expressions are indexicals functioning to refer to the addressee. Consider the contrast between (35a) and (35b), in a context where Jean, the speaker, is addressing his professor Charles:

(35) a. Tu es sympa

you[informal] are nice

You are nice

b. Vous êtes sympa

you[formal] are nice

You are nice

(35a) and (35b) have the same truth conditions: they both ascribe the property of being nice to the individual Charles, and are therefore true under the same conditions. But uses of "tu" express familiarity, whereas uses of "vous" express reverence or respect.

The observed contrast must come from another dimension than from truth-conditions, for (35a) and (35b) are truth-conditionally equivalent. The contrast can be handled with the notion of *use-conditions*. Under such an analysis, "tu" and "vous" have the same character/truth-conditions, but distinct conditions of use. Roughly, "tu" has the conditions of use sketched in (36) and "vous" the ones in (37).

(36) Use the expression "tu" to refer to your addressee only if conditions F (familiarity etc.) obtain in the utterance context.

(37) Use the expression "vous" to refer to your addressee only if conditions F' (unfamiliarity etc.) obtain in the utterance context.

Since (36) and (37) are paraphrases of a convention governing the use of certain French expressions, a competent use of "tu"/"vous" shows that the speaker has the belief that conditions F/F' obtain in the context of the conversation.

A speaker uttering (35a), if she is competent, must believe that conditions (36) obtain, and a speaker uttering (35b), if competent, must believe that conditions (37) obtain. Hearers can therefore infer that competent speakers take such preconditions to be fulfilled. As a consequence, (35a) suggests familiarity and (35b) unfamiliarity.

This is also an instance of projective-filtering content, because just like Grice's speaker's meaning - or Recanati's pragmatic implication -, use-conditions can be understood as preconditions of an action, with the difference that these preconditions are somewhat conventionalized.

That one must not be mute in order to perform an utterance is a physical necessity. That one must believe familiarity is met in order to utter "tu" is a linguistic convention. An utterance of (35a) shows that the speaker has a well-functioning vocal organs because that is a physical necessity, it might show that Jean is from Italy because he has the typical accent, and it shows that he thinks he is in a context of familiarity.

None of these inferences is rightly handled in terms of truth conditions. All the three are preconditions, and display projective-filtering behavior. For the same reason, a third and last example of projection-filtering content is the case of indexicals.

A proper name need not be uttered to have a reference ("Socrates" denotes Socrates, be it uttered of not[9]), but an indexical like "I" secures referent only when it is uttered ("I" denotes x, where x is the speaker who uttered the indexical). As a consequence, just like above, it is from the very act of utterance that hearers can infer who or what the speaker is referring to when she uses indexicals, rather than from the conventional content of the expressions that are used. This inference is therefore also a case of projective-filtering content.

Apart from all projective-filtering content which are not part of the truth-conditions or from the relevant dimension of conventional linguistic meaning, we can distinguish a class of inferences that project, even though they are truth-conditionally encoded in the meaning of the triggering expressions, because they are signaled as not being the main point of the utterance. That is my category of projective-layering content.

The idea behind this category is that propositional content can come in layers, and that only the main layer, the main point of the utterance, can be successfully affected by the insertion of compositional operators. The paradigmatic case I have in mind is the case of supplements.

The idea that propositional content comes in layers is not new. Grice first made use of the idea second-order speech-acts:

> […] the vital clue here is, I suggest, that speakers may be at one and the same time engaged in performing speech-acts at different but related levels. (Grice, 1989, p. 122)

> […] at the same time as he [a speaker] is performing these speech-acts he is also performing a higher-order speech-act of commenting in a certain way on the lower-order speech-acts. He is *contrasting* in some way the performance of some of these lower-order speech-acts with others, and he signals his performance of this higher-order speech-act in his use of the embedded enclitic phrase, "on the other hand". (Grice, 1989, p. 362)

It is important to remark that higher-order speech acts still express truth-conditional content, still talk about objects and properties. Supplements like non-restrictive relative clauses are whispered aside, but it is still clear truth-conditional content that is whispered aside.

[9] This is debated. See e. g. Pelczar and Rainsbury (1998) for discussion.

That is very different from e. g. expressive intensifiers like "damn" or "fucking" which also project but whose truth-conditional nature is, at least intuitively, very uncertain. I do not see exactly what objects and properties are added when "damn" is added in an utterance.

So even though both NRRs and expressives are insensitive to negation, they seem to be such for very different reasons. Under this second class of truth-conditional content which comes in layer, or which is "whispered aside", I propose to situate supplements, adjectival modification[10], some presuppositions, and perhaps conventional implicatures. Assimilating the expressive content of STs to this kind of projective-layering content gives rise to other candidate views, like presuppositional accounts or accounts in terms of conventional implicatures, that I discuss in the next two chapters.

Finally, take expressives like "damn", "bastard", "fucking", maybe conventional implicatures such as "but" or "even". It is not totally clear that these are encoded under the form of truth-conditions.

As Potts (2007) notices, speakers are never fully satisfied with attempts to paraphrase them[11]. Their expressivity must be conventionally encoded, but it is not clear whether we can describe it in terms of truth-conditions. I am not saying that expressivity constitutes a distinct *sui generis* category of conventional meaning, I'm saying that there is a class of similar projective meanings that are not satisfyingly describable in terms of truth-conditional layering, nor in terms of preconditions of the utterance (filtering).

Maybe in the end, we can reduce such expressives, or a subclass of expressives, to presuppositions of a sort, or to use-conditions. But as long as the possibility of such a reduction is not proven, I will call this category of non-conceptual projective content *Projective-expressive* content. The goal of the present investigation of STs and their feature of projection will then be to account for their projective-expressive potential.

In the following two chapters, I will present and critically examine different candidate accounts of the projection of STs. I will consider whether STs can be identified with i) presuppositions (a kind of projective-filtering content), with ii) conventional implicatures

[10] Note that "The big table is not red" triggers the inference that the table is big.

[11] That is the feature of expressives that Potts (2007) coined "*ineffability*".

(another kind of projective-filtering content), and then with iii) a certain sort of manner implicatures (a kind of projective-layering content).

We will eventually see that none of these hybrid linguistic accounts of STs is fully satisfactory, and will then be led to explore the psychological/mental aspect of the phenomenon.

Chapter 2. Presuppositional Accounts of Slurring Terms

The present chapter[12] explores and objects to a reduction of expressivity to presuppositional content, in particular Schlenker's *indexical* and *attitudinal* attempt at doing so (Schlenker 2007).

I start by pursuing the thoughts on hybrid meaning that I sketched in the previous chapter, and discuss the possibility of reducing all kinds of hybridity to the notion of presupposition. Although such a reduction might be tempting, I will show data displaying differences between presuppositions and other kinds of projective entities considered above, such as supplements, effects of focus, gestures, or expressives. This will pave the way for a refutation of the view that the expressive meaning of STs is of a presuppositional nature.

The main claim of the chapter is that the projective content associated with slurring terms that are embedded under filters projects more broadly than the projective content of (at least some) presuppositions under filters.

First, I show that providing such evidence requires controlling for confounds, namely *ignorance implicatures* and *intensionality*.

Second, I provide pairs of examples (namely (81)-(82)-(83)/(84) and (92)-(93)/(94)) showing that, once these confounds are controlled for, the projective content associated with STs embedded under filters projects more robustly than the projective content of (at least some) presuppositions.

[12] This chapter owes a lot to many exchanges with Benjamin Spector, to whom I express my gratitude. I also thank François Recanati, Philippe Schlenker, Yael Sharvit, Judith Tonhauser, and three anonymous reviewers, whose diverse remarks and comments greatly benefited to this work.

Such a contrastive projection behavior between standard presuppositions and STs will lead us to abandon the view that STs function the way they function because they are presupposition triggers, and to investigate other hybrid views in the next chapter.

2.1. "Presuppositionalism"

Projection is widely regarded as a test for presupposition (e. g. Levinson (1983), Soames (1989), Chierchia and McConnell-Ginet (2000), Kadmon (2001), Simons (2006), Huang (2007), to name just a few). I argue here, on the basis of the threefold distinction I established in the previous chapter, that the class of projective content is broader than that of presuppositions. This will provide first grounds for a later rejection of presuppositional accounts of STs.

Presuppositions are typically conceived of as constraints imposed by *presupposition triggers* on the conversational background. The so-called "semantic" view of presuppositions typically uses the symbol "#" to mark utterances performed in contexts where the constraints imposed on the conversational background by presuppositional material are not (and cannot be, as we will see when discussing accommodation) satisfied.

Speakers are typically reluctant to ascribe a truth-value to such deviant utterances. Since presuppositions project, and utterances triggering projective meanings in general give rise to the same sort of reluctance to ascribe a truth-value when certain conditions are met, trying to reduce all instances of projective meanings to the notion of presupposition constitutes a natural research project.

Can such a reduction be performed? Would it be progress? And indeed, wasn't the very notion of *projection* developed as a defining feature of presuppositions in the first place?

Here is an attempt at distinguishing some of the projective constructions we considered from presuppositions, at least as presuppositions are standardly conceived of. Let us first consider the case of implicatures. Can one conceive of presuppositions in terms of implicatures (or the other way around), or are the two phenomena fundamentally different in nature?

If implicatures involve a process of systematic pragmatic enrichment of literal content, can presuppositions be analyzed similarly? Or differently, if presuppositions are constraints imposed on the conversational background, can implicatures be modeled within this framework as well?

Such questions are at the core of several formal and experimental linguistics research projects (see e. g. Chemla 2009a, Egré & Magri 2008). We could for instance postulate scales like < didn't use to eat meat, stopped eating meat >, and derive the presupposed content triggered by "stop" along the lines of the following neo-Gricean mechanism, with "K", a belief-like operator (see e. g. Spector 2003, or Sauerland 2004):

(38) John stopped eating meat.

- Alternative: John didn't use to eat meat. (stronger than (38))

- Application of the maxim of Quality: K(John stopped eating meat.)

- Primary Implicature: ¬K(John didn't use to eat meat.)

- Secondary Implicature: K¬(John didn't use to eat meat.)

i. e. John used to eat meat.

As a result, an utterance of (38) would implicate that John used to eat meat. The presupposition of "stop" thus has an implicature-like treatment with the hypothesis that there are contextual scales of the kind evoked.

But it appears that there is in fact a major difference between presuppositions and implicatures. Under the scope of the negative quantifier "no", we can derive universal inferences for presuppositions, not for implicatures (Chemla, 2009b). The presence of universal inferences can therefore be seen as a criterion distinguishing between presuppositions and implicatures.

The evidence can be found for example in Chemla (2009b), first with the factive presupposition of the trigger "know" under the negative quantifier "no" (without restriction of domain, importantly):

(39) No student knows that he's lucky.

Universal presupposition: Every student is lucky.

Existential presupposition: At least one student is lucky.

The author conducted different experiments showing convincingly that presuppositions triggered from the scope of the quantifier "No" are universal. (Chemla, 2009b)

But consider now the implicature of a strong scalar item in a downward entailing environment:

(40) No student read all books.

- Alternative: No student read some books.

- Scalar inference: At least one student read some books.

Universal inference: Every student read some books.

Existential inference: At least one student read some books.

The observed inference, the inference predicted by neo-Griceans accounts (here "scalar inference") corresponds to the existential inference, not to the universal inference, as is the case with presuppositions.

Presuppositions and implicatures thus display crucially distinct behaviors in specific constructions, and are thus better kept under distinct categories of (projective) meaning, independent of the specificities of their respective analysis.

Let us now consider the case of supplements, and their difference(s) with presuppositions. Several differences can be tracked in the literature between the projective behavior of presuppositions and the one of supplements.

First of all, Potts (2005) remarks that supplements must be *non-trivial*, that is, that they have to be informative enough to be felicitously uttered, whereas trivial presuppositions are usually non-problematic.

Consider the following contrast, in a context where participants to the conversation are wondering why Mary ran out:

(41) ?Mary, who ran out, felt a bit ill yesterday.

(42) Paul ran out too.

The non-restrictive relative clause "who ran out" in (41), as well as the presupposition trigger "too" in (42), trigger the inference that Mary ran out, and as we saw, these inferences display behavior of projection.

Given the context in which these utterances are made, that Mary ran out is trivial: it is already known by the speech-participants. Crucially, (41) is judged deviant by speakers, whereas (42) is not. This contrast is taken by Potts to show that supplements must satisfy a constraint of *informativeness* that presuppositions can ignore. Supplements and presuppositions must therefore be distinct types[13].

If what is presupposed must be common ground, then presuppositions can't introduce novelty[14]. There thus seems to be a crucial difference between presuppositions, as constraint on the conversational background, and supplements: supplements can introduce novel information.

Again, in a context where my interlocutor knows Mary but does not know that she is Joe's daughter, it is felicitous to utter (43) in order to inform the addressee:

(43) Mary, who is Joe's daughter, will come to the party.

And by contrast, I can't felicitously utter: "The king of France is bald" if it is not common knowledge that France is a Monarchy.

An objection one could raise against this contrast between presuppositions and supplements is that there *are* in fact presuppositions introducing novelty. Consider a simple case where I utter "My sister stopped eating meat" in a context where participants did not now prior to my utterance that I had a sister nor that my sister used to eat meat.

[13] Note though that there might be exceptions to this observation. Schlenker (p.c.) directed my attention to examples like (i), where something like an effect of relevance intervenes, to the effect that the presupposition is not clearly *trivial*:

(i) John refuses to travel with his Lebanese wife in Israel.

[14] I neglect the phenomenon of accommodation, for the sake of clarity, but I introduce it as a counterexample just below.

Usually, such an utterance will not give rise to a presupposition failure. Participants will instead add the presupposed content to the common ground, so that we have a clear case of a presupposition introducing novel information. How can a presupposition introduce novel information if a presupposition is defined as what must already be common background?

To address this difficulty, theories of *accommodation* were developed (Lewis 1979, Stalnaker 1978; see also Thomason 1990), according to which a presupposition can be conceived of as an invitation to incorporate a certain propositional content into the conversational background.

For instance, since the determiner "the" carries a presupposition imposing a constraint on the common ground, speakers can sometimes use "the" *as if* the presupposition was common ground even when it is not, so that when the right conditions are met, participants to the conversation can charitably update the conversational background to save the speaker's utterance from oddness. This is what happens when someone says that "The king of France was beheaded in 1793" to an audience who didn't know before the utterance that France was a monarchy as that time.

Another option available to participants is simply to stop the speaker by uttering something like "wait a minute, was France really a monarchy at that time?". But letting the presupposition go through often makes the conversation smoother and the participants more cooperative. The accommodation of a certain presupposition goes through when the use of a trigger stays unchallenged by participants of the conversation.

The rationale behind the story is that a speaker can, in some cases, *do as if* a certain content was presupposed by the use of a certain trigger, knowing that cooperative participants of the conversation will add the presupposed content to the conversational background. It follows that challenging a presupposition is an uncooperative conversational move. Thus, according to theories of accommodation, presuppositions can, in some cases, introduce novel information, just like supplements.

But still, maybe the difference between presuppositions and supplements then has to do with the epistemic status of the inference they trigger. Tonhauser et al. (2013) coin this dimension the *Contextual Felicity Constraint*. The authors remark that accomodation is rare and that most presuppositions must be common knowledge *prior* to the utterance of their triggers, whereas supplements can way more easily become common knowledge *after* the utterance of their triggers (see also Schlenker, 2013).

This contrast has to do with the so-called "anaphoric" dimension of presuppositions. It is commonly held that what is presupposed must already be salient in the conversational context, and that this is not the case of supplements. Consider for instance these utterances, performed in a context where participants never mentioned anyone who arrived late:

(44) Mary, who arrived late, is a nice person.

(45) #Mary arrived late too.

In (45), the presupposition trigger "too" is said to be *anaphoric*, that is, the presuppositional content it triggers must be *antecedent* in the discourse. In (44), no antecedent is needed to resolve the supplement.

This property seems to be the flipside of that of novelty: it is because the supplement in (44) is informative that its content can be added to the conversational background without giving rise to deviance, whereas one does not see at all what content could be added to the conversational background of (45) to prevent it from being deviant[15].

But even if presuppositions and supplements don't essentially differ with regard to novelty or anaphoricity, there are other differences between the two. We can find for example a contrastive behavior under some standard presuppositional filters. Consider the case of conditionals:

(46) If France is a monarchy, then the king of France is bald.

(47) ?If Napoleon is the French emperor, then Napoleon, the French emperor, won the battle.

[15] I want to remark that there seems to be a tension between the concepts of *anaphora resolution* and that of *accomodation*, in that the concept of accommodation seems to be tailor-made to assimilate counter-examples to anaphora resolution: whenever the constraint imposed by a trigger is satisfied in a case where resolution is impossible, *accommodation* occurred.

As it is well known, the presupposition that France is a monarchy does not project to the matrix in (46). It is less clear in (47), where speakers easily interpret the NRR "the French emperor" in the consequent as a commitment of the speaker. There is a perceived tension between the antecedent and the consequent of the conditional in (47), which is absent in (46).

The same contrast is observed under disjunctive presuppositional filters:

(48) Either France is not a monarchy, or the king of France is bald.

(49) ?Either Napoleon is not the French emperor, or Napoleon, the French emperor, won the battle.

Here again, we observe that the presupposition that France is a monarchy is filtered in (48), but that (49) is odd. I will make use of these filters again when I will contrast presuppositions and STs in the next sections.

Let us now briefly consider contrasts between presuppositions and other categories of projective entities briefly presented above, such effects of focus, gestures, or expressives. Schwarzschild (1999) proposes the following analysis of the projective meaning triggered by focus:

(50) John bought a RED car.

a) Given (*not* presupposed): there is an *m* such that *x* buys a car *m*.

b) If the discourse does not satisfy a), then we suppose that a) is presupposed.

That something being "given" is different from being "presupposed" can be shown with the following contrast between (51) and (52):

(51) John denies that Paul bought a blue car; did Paul buy a RED car?

(52) #John denies that Paul came; did Mary come too?

This contrast can also be observed under disjunctive filters:

(53) Either Paul didn't buy a blue car, or he bought a RED car.

(54) ?Either Paul didn't come, or Mary came too

There might also exist differences between presuppositions and gestures (Schlenker, 2014, Ebert et al, 2011). Gestures seem to trigger what Schlenker (2015), coins "cosuppositions", which is to say, they trigger presuppositions that are conditional on the assertive part.

The cosupposition of (30a) for instance is only that, *if* Luca were to help his son, it would be by pushing him upward. The debate is not settled, Eber et al. (2011) claim to the contrary that gestures in fact *are* supplements.

Finally, attitudinal expressions, or expressives, are intuitively different from supplements and from presuppositions. Supplements are truth-conditional and objective, they are about properties that objects have or do not have, whereas what seems to be specific of attitudinal expressions like "fucking" is their intrinsically subjective dimension.

Expressives also seem distinct from presuppositions with regard to their behavior under plugs (although it might depend on the specificities of the propositional content we attribute to STs), as well as with regard to their pragmatic effects (Schlenker 2007, Richard, 2008). I now will develop a detailed argument in favor of this hypothesis.

To sum up, although it might be tempting to try a reduction of all kinds of projective content to presupposition, it seems that there are differences between presuppositions and other kinds of projective entities, such as supplements, effects of focus, gestures, or expressives. I now turn to a refutation of the view that the expressive meaning of STs is of a presuppositional nature.

2.2. Introduction to a Presuppositional Account of Slurring Terms

We can extract from my above list of explananda that any adequate theory of STs has three major tasks. An adequate theory should first account for the *extension* of STs, in order to be able to derive the truth-conditions of slurring sentences. Is the extension of a ST the same as the one of its neutral counterpart (NC), or is it empty, as some have suggested (Hom and May, 2013)? If the former, how do STs differ from their neutral counterpart?

Second, the theory must account for the *expressivity* of STs, and has to show explicitly how they come to have the powers they have. If the derogatory attitudes of speakers belong to the meanings of STs, how come a competent speaker may grasp the meaning of a ST without sharing the attitude?

And third, the account must be able to adequately predict the *projective* profile of STs under all kinds of constructions.

In dealing with these three basic questions, the theory should account for the different distinctive characteristics of STs - their offensiveness and other illocutionary forces, the unwillingness of speakers to evaluate the truth or falsity of slurring sentences, the autonomy of ST's derogatory force from the speaker's mental states, and so on and so forth.

It should thus explain why competent speakers know the meaning of STs but refuse to assert slurring sentences. It should also take a stand on which module of our cognitive abilities - syntax, semantics, pragmatics, social competences etc. - is responsible for each of these peculiar properties.

Facing these questions, we saw that the natural move was to propose a multi-dimensional account of STs. Indeed, it seems that we can replace the ST of a slurring sentence by its NC without affecting its truth-conditional content: the two resulting sentences express the same proposition, despite our intuitive judgments, which might be explained by other means.

Thus, in virtue of a descriptive component, a ST seems to denote the same individual or group of individuals as its neutral counterpart. That is what I coined the "co-extensionality thesis" (CET). But we also have to account for the peculiar features of STs that make them different from their neutral counterpart: their offensiveness, our reluctance to assent truth to slurring sentences and so on; these properties are often gathered up under the idea of an additional expressive component of STs.

As we just saw, most linguists and philosophers have been attracted by Hybrid Expressivists Accounts (HEA), that is with an intuitive distinction between a descriptive component of STs, responsible for their classificatory power, and an expressive component of STs, responsible for their attitudinal manifestations.

There are several ways to analyze STs as terms with a hybrid descriptive/expressive content whose parts are separable. Indeed, as soon as hybridity is taken on board, it becomes possible to rely on any of the numerous distinctions between kinds of meanings that were drawn for different purposes in the last fifty years or so. Here are some of the main theoretical distinctions that could in principle be used to model STs:

- Reference vs. Mode of presentation
- Sense vs. Tone
- Content vs. Character
- Content vs. Force
- At-issue/truth-conditional content vs. Presuppositional content
- At-issue/truth-conditional content vs. Conventional implicatures
- At-issue/truth-conditional content vs. Conversational implicatures

I identify the main four families of hybrid theories of STs: Presuppositional accounts (Macià 2002, 2006, Sauerland 2007, Schlenker 2003, 2007, 2014), Conventional Implicature accounts (Potts 2007, McCready 2010, Gutzmann 2015) or use-conditional accounts (Kaplan 1999, Predelli 2013), and speech-act accounts (Nunberg 2017, Bianchi 2014, Langton et al. 2012, Langton 2012). The remaining of the present chapter focuses on presuppositional accounts.

According to Potts and others, expressives in general and STS in particular display peculiar properties that require the introduction of a novel dimension of meaning, independent of other kinds of content (Potts 2007, McCready 2010, Gutzmann 2015).

On the other hand, other authors have suggested more parsimoniously that provided certain extra features are acknowledged, the behavior of STs could be handled in a standard presuppositional framework, with no need for postulating an additional dimension in one's theories of meaning on top of whatever is independently needed to take care of presuppositions (Macià 2006, Sauerland 2007, Schlenker 2003, 2007, 2016, Cepollaro and Stojanovic 2016).

In particular, Schlenker (2007), responding to Potts (2007), proposes that expressives carry a presupposition that is *indexical* (evaluated w.r.t. a context), and *attitudinal* (predicates something of the agent's mental state)[16].

The question of whether STs can be analyzed as presuppositions has received both positive (Macià 2006, Sauerland 2007, Schlenker 2003, 2007, 2016, Cepollaro & Stojanovic 2016) and negative (Potts 2007, Richard 2008, Davis and McCready 2016) answers in the literature.

The data I discuss below supports the negative side, which could constitute positive evidence in favor of use-conditional accounts of STs and expressivity *à la* Potts for instance, or maybe in favor of other types of accounts such as Nunberg's Gricean view (Nunberg 2016, see chapter 3), or Hom and May's radical truth-conditional theory (Hom and May, 2013, 2015, see chapter 6).

The idea that STs carry an expressive presupposition is a straightforward possibility to account for the fact that expressive content projects out of various embeddings like negation (55), conditionals (56), modals (57), questions (58) and so on:

- *Context*: Salma was never a meat eater; none of the participants to the conversation are prejudiced against German people in any way whatsoever[17].

[16] Incidentally, anticipating the potential existence of various perspectival readings, Schlenker also defines such expressive presuppositions as being sometimes *shiftable* (i. e. the context of evaluation need not be the context of the actual utterance). This property will not be relevant in what follows.

[17] All the data I will discuss are introduced with a context so as to be able to perceive projecting inferences. They were all first judged introspectively in French, then confirmed in

(55) a. #Salma didn't stop eating meat.

b. !Salma didn't marry a boche.

Because of the presupposition trigger "stop" in (55a), the utterance triggers an inference that Salma used to eat meat. That inference conflicts with what we assumed was common knowledge in the context, and the utterance is thus perceived as odd[18].

Similarly (same context as in (55)):

(56) a. #If Salma stopped eating meat, then her mother is happy.

b. !If Salma married a boche, then her mother is happy.

(57) a. #Salma might have stopped eating meat.

b. !Salma might have married a boche.

(58) a. #Did Salma stop eating meat?

b. !Did Salma marry a boche?

The a-sentences illustrate the familiar fact that presuppositional content projects out of negation, if-clauses, or questions: all the a-sentences license the inference that Salma used to eat meat. The b-sentences illustrate the fact that the expressive content of STs (and other expressives) have a projection behavior similar to that of presuppositions. An utterance of

conversations with peers, and presented in a semi-formal setting to three native speakers of English. Participants were orally exposed to the conversation contexts shown in the chapter, asked to imagine a certain utterance to be performed in that context, and then asked to judge the extent to which they could infer that the speaker was racist or prejudiced.

[18] Recall that in order to stay neutral with regard to the discussed reduction of expressivity to presuppositional content, I use the symbol "!" to mark the expressivity of utterances such as (55b), and in particular, the presence of an inference about the speaker's emotional state that conflicts with the specified context. I use the symbol "#" for plain vanilla presupposition failures.

any of the b-sentences licenses the inference that the author of the utterance is prejudiced against German people.

This parallelism, *prima facie*, could motivate attempts at reducing expressive to presuppositional content. And the perspective of not having to unnecessarily posit a new dimension to our general view of meaning (in contrast with Potts approach) is indeed appealing.

The main claim of the chapter is that the projective content associated with STs that are embedded under filters projects more broadly than the projective content of (at least some) presuppositions under filters. This claim is made in two steps. First, I show that providing such evidence requires controlling for confounds, namely *ignorance implicatures* and *intensionality*.

Second, I provide pairs of examples (namely (81)-(82)-(83)/(84) and (92)-(93)/(94)) showing that, once these confounds are controlled for, the projective content associated with STs embedded under filters projects more robustly than the projective content of (at least some) presuppositions.

At least two versions of a presuppositional theory might account for the behavior of STs (Richard 2008). A first one could hold that STs carry a presupposition in virtue of which, if it is not already in the conversational background, the utterance is a presupposition failure.

However, that a presupposition is not already in the common background doesn't always lead to presuppositional failure, as we saw (see also Yablo's (2006) notion of non-catastrophic presupposition failures). Participants to the conversation who did not hold the belief that Salma used to eat meat could well rescue (55a) in taking it to indicate, in parallel to the main proposition, that Salma was a meat eater (that is *accommodation*).

Applied to STs, this mechanism would lead us to hold that an utterance of "John is a boche" somewhat invites participants to accommodate, that is, to take for granted a racist propositional content that was initially absent from the conversational background. Both mechanisms (presupposition failure and presupposition accommodation) could coexist, and are to a large extent independent of the specific content we ascribe to the alleged expressive presupposition.

In what follows, I focus on Schlenker's (2007) version of a presuppositionnal account of STs, because it is to the best of my knowledge the presuppositional account that is the most likely to derive as wide a projection profile as is needed, in virtue of two additional features: *indexicality* and *attitudinality*.

Schlenker provides the following lexical entry for the ST "honky" (see Kaplan 2001), with respect to a context (c) and a world (w):

(59) [[honky(c)(w)]] ≠ # iff the agent of c believes in the world of c that white people are despicable. If ≠ #, [[honky]](c)(w) = [[white]](c)(w)

According to this analysis, "honky" and "white" or "white person" make the same truth-conditional contributions to utterances in which they appear, and differ with regard to their presuppositional import.

Where "white" does not trigger any presupposition, (or at least no presupposition that is relevant to the present discussion) "honky" triggers a presupposition of a particular sort. The presupposition is about the agent of the context (it is indexical), and more specifically, it is about the agent's attitudes (it is attitudinal).

According to Schlenker, these linguistic properties are sufficient to derive the effects of STs. The indexical character of the presupposition, together with the assumption that there are shiftable indexicals (Schlenker 2003, Sauerland 2007) would yield the dependency to a particular perspective that STs and other expressives have been noted to display (Potts 2007). Moreover, the presupposition of STs would be automatically accommodated, because subjects are usually taken to be authoritative on their own attitudes.

And according to such a presuppositional view, expressives like "honky" are predicted to follow the same patterns of projection as what is expected from a presupposition that is indexical and attitudinal.

I now turn to an attempt at falsifying this prediction. As the realm of presuppositions is quite diverse and heterogeneous when it comes to projection, I will systematically compare the behavior of STs to that of both soft triggers (e. g. the factive "know", or the existence presupposition triggered by "the") and hard triggers (e. g. "too"), whose import is very difficult, not to say impossible, to accommodate.

2.3. Projection Under Simple Filters

It appears that the expressive content of STs projects to the matrix position even in environments where standard presuppositions tend to get filtered. Although presupposition failures usually arise even when the trigger of a presupposition that is not satisfied in the context of utterance is embedded, we briefly saw in the section 2.1 that there are linguistic environments in which presuppositional material interacts with the standard descriptive dimension, to the effect that projection is blocked (the so-called presupposition filters, Karttunen 1973).

I begin with the comparison between STs and three standard presupposition triggers ("the", "know", and "too") under two such filtering environments (disjunctive filters and conditional filters). Facing the need to control for confounds, such as ignorance implicatures and intensionality, I will then consider two more adapted filtering environments.

2.3.1. Disjunctive Filters

Let us start with the simple sentences (60)-(64) to construct the disjunctive filters (65)-(68), following Schlenker's (2007) discussion:

- *Context* (Take this context to be the default context in which to evaluate all data without context in the remaining of the chapter): none of the participants to the conversation are prejudiced against German people in any way whatsoever and France is not a monarchy.

(60) !John is a boche.

(61) #The monarch of France is bald.

(62) #My colleagues are Germanophobic too.

(63) I am Germanophobic.

(64) France is a monarchy.

Adapting Schlenker's lexical entry, we have $[[(60)]](c)(w) = \#$ if s(c) isn't Germanophobic, else $[[(60)]](c)(w) = 1$ iff John is German, with s(c) = the speaker of the context.

Similarly, (61) presupposes that there exists a (unique) king of France and (62) that someone salient other than the speaker's colleagues is Germanophobic. Trivially, the truth-conditions of (63) are such that $[[(9)]](c)(w) = 1$ iff s(c) is Germanophobic, and (64) is true, relative to a context *c* and a world *w*, if and only if France is a monarchy in *c*.

Building these blocks together, consider the following contrast:

(65) France is not a monarchy, or the monarch of France is bald. $((\neg (64)) \vee (61))$

(66) ?I am not Germanophobic, or my colleagues know that I am. (adapted from Schlenker 2016, p. 47)

(67) ?I am not Germanophobic, or my colleagues are Germanophobic too. $((\neg (63)) \vee (62))$

(68) !I am not Germanophobic, or John is a boche[19]. $((\neg(63)) \vee (60))$

As we see in (65), the presupposition that there exists a (unique) monarch in France can be locally accommodated - that is, roughly its content does not systematically project to the matrix and can stay stuck in the embedded phrase, here the second disjunct (see Karttunen 1974, Heim 1983) -, when it is negated in the first disjunct: the presupposition is not inherited by the entire utterance.

But in (68), the presupposition that the speaker is Germanophobic seems to be inherited by the whole utterance (or at least, the sentence is very odd), even though it is negated in the first disjunct. (66) and (67) are somewhat less clear, we will understand why in what follows.

Considering only (65) and (68) we observe a first contrast between the expressivity of STs and presuppositions in a disjunctive filter, contrary to what a presuppositional view of STs would expect[20]. Can a presuppositional analysis explain this contrast away?

[19] Schlenker constructs and discusses a similar example to address precisely that difficulty.

Note that some previous work has noted that expressives can in fact be evaluated just in their embedded position (e. g. Kratzer 1999, Schlenker 2003, Potts 2007). Consider for example (69):

(69) Every member of my family is so racist. I hate it that they won't accept that I married a white person. It's so embarrassing that everyone in my family thinks I married a honky[21].

Naturally, it may be argued, we understand the speaker of (69) as not sharing her family's racism at all. However, such considerations do not speak against the general observation that there is a clear preference for the projective reading.

First, embeddings under attitude verbs are trickier to interpret than it might first appear, because of the potential intervention of perspectival operations that are to the best of my knowledge not yet well understood.

[20] Potts makes the observation that expressives display different projection behaviors under the scope of propositional attitude predicates, whereas presuppositions are typically cancelled in these environments (Potts 2007):

(i) !Mary believes that Paul realizes that honkies are tall.

(ii) Mary believes that Paul realizes that there is a queen in Germany.

Potts argues that under one reading of (i), the presupposition engendered by "realize" is satisfied if Mary believes that there is a queen in Germany, even if there is no queen in Germany, whereas every reading of (ii) is offensive (that is, one could not evaluate the offensive presupposition as embedded, it is always evaluated in the matrix).

But as an anonymous reviewer of this chapter (submitted as a paper in *Semantics and Pragmatics*) rightly remarked, this apparent contrast is not telling because it is in fact not the case that presuppositions are typically cancelled under attitude verbs. The consensus seems to be that presuppositions triggered under attitudes are evaluated in both the embedded position and in matrix position (e. g. Heim 1992, Zeevat 1992, Geurts 1999, Singh 2008, Beaver and Geurts 2011, Schlenker 2011).

[21] I thank an anonymous reviewer for bringing this case to my attention.

Second, even if expressives are in some cases evaluated only in their embedded position, there is still an overwhelming preference for the matrix position. That is in itself puzzling, even granting that there is some indexical component in the presuppositional content of STs.

As Schlenker notices, his analysis has in fact the resources to explain the contrast observed under disjunctive filters. In virtue of the indexical nature of the presupposition that Schlenker posits for expressives, an oddity might arise in (69), as well as in (66)-(67), and not (65) through the (possibly obligatory) parallel computation of ignorance implicatures: whereas it is conceivable that the speaker does not know whether or not France is a monarchy, it is hard to buy that she does not have access to her own attitudes.

As a result, ignorance implicatures are non-problematically derived in disjunctive statements like (65), but they clash with common world-knowledge in disjunctive statements like (68) (see e. g. Magri 2009 for more on that point).

Qua disjunctive statements, utterances of (65)-(68) will undergo the following neo-Gricean enrichment, following Sauerland's (2004) proposal for the computation of scalar implicatures:

(70) Take A and B, two propositions, and K, an unspecified epistemic operator:

- We assume that <$A \wedge B$, A, B, $A \vee B$> form a scale - Utterance: $A \vee B$

- Application of the maxim of Quality: $K(A \vee B)$

- Generation of alternatives: A, B, $A \wedge B$

- Primary implicatures[22]: $\neg(K(A))$; $\neg(K(B))$; $\neg(K(A \wedge B))$

- Secondary implicatures[23]: $\neg(K(\neg(A)))$; $\neg(K(\neg(B)))$

[22] Given that each of the three alternatives asymmetrically entail the utterance, they are more informative than the utterance

[23] Entailed by $K(A \vee B) \wedge \neg(K(A)) \wedge \neg(K(B)$. Intuitively, the speaker cannot believe that A is false, as given $K(A \vee B)$, she would thus believe that B is true; but by NEG(K(B)) she

Taken together, disjunctive statements of the form $(A \vee B)$ trigger the inference that $((\neg(K(A))) \wedge (\neg(K(\neg(K(A))))))$, or in words, that the speaker has no belief about whether A is the case or not the case. That is an ignorance implicature.

In the case of (65), this will give rise to the inference that the speaker does not know whether France is a monarchy or not, which is acceptable. But in the case of (66)-(68), this mechanism of enrichment gives rise to the inference that the speaker does not know whether she herself is Germanophobic or not. That implicature, plus the common world-knowledge that one's own attitudes are transparent, correctly predicts oddity for (66)-(68) and felicity for (65).

Comparing (68) and (66) for instance, we see how the indexical character of the presupposition, plus the impossibility of deriving ignorance implicatures when speakers talk about their own attitudes, could deal with the objection of contrastive behaviors under disjunctive filters. In order to control for the confounding factor of ignorance implicatures, one shall therefore test presupposition filters that trigger the right inferential mechanisms and do not give rise to oddity, even in the presence of indexicality.

2.3.2. Subjunctive Conditional Filters

In order to control for ignorance implicatures in comparing STs with other presuppositions under filters, I now compare the behavior of STs and that of presuppositions under a presupposition filter where ignorance implicatures do not interfere. Subjunctive conditional constructions that display in the antecedent the content of a presupposition triggered in the consequent seem to constitute such a case.

Compare the following conditional statements:

(71) If France was a monarchy, the monarch of France would be bald.

does not believe that B is true. The converse entails that the speaker does not believe B to be false.

(72) If I were Germanophobic, then my colleagues would know that I am.

(73) If I were Germanophobic, my colleagues would be Germanophobic too.

(74) !If I were Germanophobic, then John would be a boche.

Given that i) ignorance implicatures do not interfere in such conditionals - as shown by the acceptability of e. g. (72) -, ii) the presupposition that France is a monarchy is filtered out in (71) and the presupposition that the speaker is Germanophobic is filtered out in (72) and (73), iii) the racist expressive content is not filtered out in (74), it appears that the contrast between STs like "boche" and standard presupposition triggers restores the initial filtering problem that presuppositional views of STs faced, even with indexical and attitudinal presuppositions *à la* Schlenker.

But on closer inspection, we notice that, again, the indexical character of the expressive presupposition, on a par with a counterfactual analysis of conditionals, could well derive the intended results. Taking R to be the relevant accessibility relation, w* the actual world, and under a dynamic strict analysis for subjunctive contitionals (von Fintel 2012), there would in fact be two alternative ways of characterizing the conditional presupposition of (74):

(75) $\forall w \in R(w^*)$ (([[(9)]](c)(w) = 1) → (the speaker of c is Germanophobic in w))

(76) $\forall w \in R(w^*)$ (([[(9)]](c)(w) = 1) → (the speaker of c is Germanophobic in w*))

That is, an utterance of (74) "If I were Germanophobic, then John would be a boche" expresses the proposition that in all (epistemically) accessible worlds where the speaker is Germanophobic, John is a German in that world (75), or alternatively, John is German in the actual world (76).

As the property of being German ascribed to John is expressed through presuppositional material (under the view that "boche" carries an expressive presupposition), the difference between the two truth-conditional analyses is a crucial matter: if it is w rather than w* that plays a role for the satisfaction of the consequent (75), then the world variable is bound by the intensional operator, and the indexical expressive presupposition will be evaluated in the worlds that are quantified over (say, in a point of evaluation w(c)), and the anti-German sentiment will be ascribed to the utterer as she would be in this hypothetical world, not to s(c), the actual speaker of the utterance. No projection of the expressive material is thus predicted here.

But if it is w* that is relevant for the evaluation of the consequent (76), then the indexical expressive presupposition will be evaluated in the world of the actual context, and the anti-German sentiment will be ascribed to s(c), the utterer of (74). So Schlenker's account can in fact predict the projection of the expressive presupposition in that case. More precisely, just like the presupposition of (71) is conditional (Schlenker 2008),[24] the presupposition of (74) that would be predicted under Schlenker's analysis is either of the following:

(77) $\forall w \in R(w^*)$ (([[(63)]](c)(w) = 1) → presup'((60))(c)(w) = 1))

(78) $\forall w \in R(w^*)$ (([[(63)]](c)(w) = 1) → presup'((60))(c)(w*) = 1))

that is,

(79) $\forall w \in R(w^*)$ (([[(63)]](c)(w) = 1) → (s(c) is Germanophobic in w))

(80) $\forall w \in R(w^*)$ (([[(63)]](c)(w) = 1) → (s(c) is Germanophobic in w*))

[24] Roughly, the presupposition of a subjunctive conditional statement is that, in all accessible worlds, if the antecedent is true in that world, then the presupposition of the consequent is satisfied in that world (which I write, for the consequent q of an utterance in c, presup'(q)(c)(w) = 1).

For conditional statements where the descriptive content of the antecedent precisely is the presuppositional content of the consequent, we obtain the presupposition that, in all accessible worlds, if France is a monarchy in these worlds, then the presupposition of "the king of France is bald" is satisfied; that is, if France is a monarchy in these worlds, then France is a Monarchy in these worlds: hence the presupposition being satisfied trivially.

In the case of an indexical presupposition, as one is now considering, things might be different as the satisfaction of the presupposition of the consequent might well be indexed on the actual world w* rather that on the point (world) of evaluation w.

In other words, an utterance of "If I were Germanophobic, then John would be a boche" is felicitous if, in all accessible worlds where the speaker is Germanophobic, the speaker is Germanophobic at that world (79) (or alternatively, in the actual world (80)).

In the case of (79), the conditional presupposition is predicted to be trivially satisfied, and consequently, an utterance of (74) is wrongly predicted to be felicitous. But if the option to have w* featuring in the computation is left open as it is the case in (78) and (80), then the conditional presupposition is not at all trivial, and imposes its non-trivial constraints on the utterance context itself: (74) is predicted to be presuppositional (in the sense that it forces hearers to accommodate the proposition that the actual speaker is prejudiced against Germans) in most contexts.

To put it in different terms, the above argument rests on the facts that, because subjunctive conditionals are intensional operators, and because Schlenker's expressive presuppositions are indexical, the contrast between (71)-(72)-(73) and (74) is useless.

If the presupposition triggered by "boche" is indexical in the sense that the French second person pronoun "tu" is, then it is indexical in a double way: i) it is about the speaker, and ii) it imposes a condition on the world of the utterance's context (not on the point (world) of evaluation).

Because Schlenker's expressive presupposition is indexical, the consequent of (74) could very well presuppose that s(c) is Germanophobic in the utterance world w*. The world-variable is not necessarily bound by the intensional operator in the case of expressive presuppositions, and the contrast between (74) and (71) could be explained away by recognizing that the presupposition of expressives like "boche" is indexical also in the sense that their content is always to be evaluated relative to the actual world.

Conditionalization wouldn't affect them, as intentional operators do not quantify over the actual world. The presupposition carried by "boche" and such, when uttered by s, will be satisfied relative to the pair <c, w>, where w is an arbitrary (accessible) world, if, and only if, the speaker s despises German people in w*. That is, the indexical presupposition carried by "boche" could impose a constraint on a parameter of the context itself, and not necessarily on a parameter of the point of evaluation.

On the other hand, the presupposition of (72) "If I were Germanophobic, then my colleagues would know that I am", when uttered by s, is satisfied in a pair <c, w> iff the speaker s is Germanophobic in the world of evaluation w(c). The consequent of (74), evaluated in a pair <c, w(c)>, must thus have a standard truth-value (not a presupposition failure) in all pairs <c, w'> such that w' satisfies the antecedent.

But for the presupposition of "boche" to be satisfied in <c, w'>, one considers only c, so that the speaker must have an anti-German sentiment in the world of the utterance's context itself for the presupposition to be met.

An utterance of "My colleagues would know that I hate German people" will be interpretable relative to a pair <c, w*> only if s(c) thinks, in w(c), that German people are worthy of contempt and so on in w(c). But an utterance of "John is a boche" is interpretable relative to a pair <c, w*> only if s(c) despises German people in w* (not in w(c)). So Schlenker's theory *does* make the correct predictions here provided i) a dynamic strict analysis for subjunctive conditionals and ii) a doubly indexical character of the attitudinal presupposition.

Taking stock, we cannot find a contrastive behavior under disjunctive filters because of an interfering pragmatic phenomenon of ignorance implicature, nor under subjunctive conditional filters because of the (potentially doubly) indexical character of Schlenker's expressive presuppositions. One shall therefore test extensional contexts - that is contexts in which the world of evaluation is not affected by the conditional and is thus the same as the world of the utterance's context - in which ignorance implicatures are controlled for.

I consider below two such cases. I first go back to disjunctive filters (as they only involve extensional operators) in an imaginary and somewhat artificial context blocking ignorance implicatures. Second, I consider the case of conjunctions under negation.

2.4. Projection Under Complex Filters

2.4.1. Disjunctive Filters Without Ignorance Implicatures

Let us go back to disjunctive filters (as it only involves extensional operators), but this time with a context in which Grice's maxim of quantity is suspended, resulting in the cancellation of ignorance implicatures. Imagine a world where Caucasians tend to be oppressed, marginalized, disenfranchised and so on.

Imagine a game show in that world, where a participant is supposed to guess the identity of her interlocutor, hidden behind a curtain. The hidden interlocutor can give hints, like "I am the son of a baker", or "Either I am a journalist, or I am the son of a baker", in order to help the candidate eliminate hypotheses and eventually narrow down her identity.

In this sort of context, "either, or" constructions do not trigger ignorance implicatures, as participants are purposefully less informative than they otherwise could be (see Fox 2014 for an earlier discussion of games in which the maxim of quantity is deactivated).

Indeed, when the hidden interlocutor utters, "Either I am a journalist, or I am the son of a baker", one does not infer that she does not know whether she is a journalist or not, one understands she is giving hints to the candidate rather than expressing her beliefs in maximizing informativity. As the maxim of quantity is suspended, ignorance implicatures will not be derived. Having set up this specific context will now allow us to test the projection behavior of STs in the right kind of environment.

- *Context*: Mary is a candidate in the game show and must guess the identity of someone hidden behind the curtain. At that stage in the game, she is hesitating between three individuals (who happen to have a daughter): Bob, an anti-Caucasian journalist whose daughter knows he is anti-Caucasian and who is anti-Caucasian too; John, an anti-Caucasian baker whose daughter does not know he is anti-Caucasian and who is not anti-Caucasian herself; and Alfred, who is notoriously not anti-Caucasian, and also has a daughter. The hidden interlocutor says: "I will give you a hint, but I shall not be too informative":

(81) Either I don't hate Caucasians, or my daughter knows I hate Caucasians.

(82) Either I don't hate Caucasians, or my daughter hates Caucasians too.

- *Context*: Similar game context, but this time with Mary hesitating between five individuals (who all have a daughter): three non-racist and two anti-Caucasian. Among the non-racists, one has a daughter who married a Caucasian, one has a daughter who did not marry a Caucasian, and we do not know about the daughter of the third. The daughter of the first racist married a Caucasian, unlike that of the other.

(83) Either I don't hate Caucasians, or I hate Caucasians and my daughter married a Caucasian.

In (81)-(82)-(83), the racist presupposition is triggered not by a ST but by the factive "know", the anaphoric "too", and mere at issue content, respectively. Plus, the negation of the racist content being investigated features in the first conjuncts.

Interestingly, they do not appear to convey any expressive or racist content. The three sentences seem to have been uttered as a mean to rule out some candidates (John, the anti-Caucasian baker whose daughter does not know he is, in (81)-(82) for instance), and thus express something like the disjunction "I am Alfred or I am Bob".

Importantly, Mary has no grounds to draw an inference that the speaker is anti-Caucasian: she still has two options to go with, Bob and Alfred, and has no grounds to distinguish between the two.

But things are different if the racist expressive content is triggered by a ST instead. Consider (84) as uttered in the same context as that of (82) or (83):

(84) !Either I don't hate Caucasians, or my daughter married a honky.

In (84), where the anti-Caucasian content of the consequent is conveyed though the use of a ST, it seems that one can legitimately infer that the speaker is prejudiced against Caucasians. Contrary to (81)-(82)-(83), Mary does have evidence after (84) that it is Bob, the anti-Caucasian journalist, talking behind the curtain, rather than Alfred the non-racist (or than John of course).

But presuppositional theories of STs do not predict any difference between (81)-(82)-(83) on the one hand and (84) on the other hand. The presupposition of honky in (84) should in fact be filtered for the same reason as the "expressive" presuppositions of "know" and "too" in (81)-(82)-(83) are filtered. The fact that expressivity projects in (84) exhibits a wrong prediction of Schlenker's and other presuppositional reductions of expressivity (in the case of STs at least).

Now surely, given that the racist expressive content in (84) projects, the speaker must be ascribed a certain degree of hatred towards Caucasians. It is unclear then what kind of an epistemic state she could be in to make such a disjunctive statement, conveying at one and the same time a negative attitudinal state and uncertainty about her being in that state. What exactly could she have intended to communicate then?

But presuppositional accounts cannot sensibly rely on such pragmatic oddity because this oddity arrives only after the tested expressive content projects, and the mere fact that expressivity projects here is sufficient to falsify presuppositional theories. In fact, the fact that (84) feels pragmatically odd is predicted only if the derogatory content projects, and is thus not predicted by the presuppositional approach.

In (84), it looks like the speaker intended to make a neutral disjunctive statement in order to eliminate hypotheses and ended up accidentally slipping a ST, thence revealing her true attitudes. But the fact that slipping a ST in that environment does in fact reveal her attitudes is evidence that the expressive content is not plugged where it was expected to be according to presuppositional accounts.

2.4.2. Paraphrases

Before drawing any conclusions from the above contrasts, we should carefully consider a potential methodological worry. Is it still possible that STs in fact do carry expressive presuppositions, but that their filtering is harder to detect because the filtering constituents I chose (e. g. "I don't hate/despise Caucasians") are unfaithful paraphrases of the actual presupposition?

If the actual expressive presupposition triggered by "honky" happened to be far richer and more complex than "I hate/despise Caucasians", then one could not expect such a simple paraphrase to be able to play the canceling role we expected it to play.

The following argument should help cast this worry aside. Let us first admit that the paraphrase I used is oversimplified and does not capture the true expressive content of STs. Let me then introduce instead p1, the actual expressive presupposition of the ST (whatever it happens to be). p1 might turn out to be as simple as "I hate Caucasians", or as rich and complex as "Caucasians are generally cruel and ought to be the target of negative moral evaluation simply in virtue of being Caucasians", or maybe even ineffable (see Potts 2007 on ineffability).

Consider now p2: "I am disposed to use the well-known racist insult to refer to Caucasians". p2 contextually entails p1: a person could not sensibly be disposed to call Caucasians "honkies" unless she is at least as much of a racist as if she had actually uttered the term "honky". That being said, imagine again that Mary is in the context of the guessing game, and consider the following pair of utterances by her hidden interlocutor:

(85) a. !Either I don't hate Caucasians, or my daughter married a honky. ((84) repeated)

b. Either I don't hate Caucasians, or my daughter married a person I would be disposed to refer to using the well-known anti-Caucasian insult.

Note that generally, utterances of the form $(\neg A \vee B)$, where B presupposes that p, presuppose $(A \rightarrow p)$. In (85a), the predicted presupposition is thus that if the speaker hates Caucasians, then the presupposition of honky (p1) is satisfied, which is almost trivial. Under the form of (86a), to compare with (86b), the presupposition does become truly trivial:

(86) a. !Either I have no disposition to use the well-known anti-Caucasian insult, or my daughter married a honky.

b. Either I have no disposition to use the well-known anti-Caucasian insult, or my daughter married a person I would be disposed to refer to using the well-known anti-Caucasian insult.

But we observe an unexpected contrast here: it seems that p1 can be inferred from a-members but not from b-members of the above data. And given that the second disjunct of

(85b)-(86b) in fact contextually entails whatever expressive content (p1) is triggered by the use of "boche" in (85a) and (86a), how could it be that p1 leaks out of the plug when triggered by "boche", but is canceled when triggered by p2?

That cannot stem from an oversimplification of the paraphrase, because it features in both members of the pair and p1 projects in only one of them[25]. So overall, from the point of view of presuppositional theories of STs, this contrast between (81)-(82)-like sentences - with a "mere" racist presupposition - and (84)-like sentences - with a ST - remains highly unexpected, and that observation is independent from the potential weaknesses of the chosen paraphrases for expressive content.

Nevertheless, the scenario I just constructed in order to test the above data might still be judged to be too artificial and unecological to allow us to draw reliable conclusions. I thus now turn to another contrast in filtering.

2.4.3. Negated Conjunctions

I present a second example of a non-intensional context displaying a contrast between the projective profile of STs and that of presuppositions, one that does not involve imaginary *scenarii*, nor conflicting ignorance inferences.

There is a specific sort of negated conjunction that works as a presuppositional filter: the negation of a conjunction displaying in the descriptive part of the first conjunct, the presuppositional import of the second conjunct. First, note that such conjunctions are not presuppositional (Karttunen 1973):

[25] Note that the judgments are less obvious with a clear mentioning case:

(i) !Either I don't hate Caucasians, or my daughter married a person I would be disposed to call a "honky".

It seems that at least some expressivity leaks out of quotation here.

(87) France is a monarchy and the monarch of France is bald.

(87) does not presuppose that France is a monarchy. Indeed, although the consequent alone does carry the presupposition that France is a monarchy, adding precisely that content in the first conjunct has the effect of restricting the context set of the evaluation of the consequent precisely to those worlds in which the presupposition of the consequent is already satisfied. No further restriction is needed (in a dynamic framework, see again Schlenker 2008), that is, (87) is not presuppositional.

Now, of course, an utterance like (87) will still convey the false information that France is a monarchy, but that is only because of its descriptive material: what is *said* is that France is a monarchy. One can therefore safely construct its negated alternative:

(88) It's completely false that France is a monarchy and that the monarch of France is bald.[26]

And as expected, the result is neither presuppositional nor does it convey the false information that France is a monarchy. Now consider the same constructions involving Germanophobic content:

(89) I am Germanophobic, and my colleagues do not know it.

In (89), the context set for the evaluation of the second conjunct is restricted to precisely those worlds which satisfy the presupposition, so that (89) is not predicted to be presuppositional. An utterance like (89) will still be very offensive in virtually any context, but again that is only because of its descriptive material: it is said that the speaker despises German people.

And since (89) expresses anti-German sentiment because of its descriptive material, one can safely construct its negated alternative and expect the result to be neutral:

(90) It's completely false that I am Germanophobic and that my colleagues do not know it.

[26] I use the unnatural "it's completely false that" form for negation in order to avoid potential complications that could arise because of the ubiquitous metalinguistic readings of standard negation, (provided that phrase actually blocks metalinguistic construals).

And indeed, (90) does not seem to commit the speaker to anti-German sentiments. So far so good. I note here that for puzzling reasons that ought to be clarified independently, (90) and other such constructions appear not to trigger ignorance implicatures, although they are equivalent to some disjunctive filters such as our earlier (65)-(67)[27].

(90) for instance is formally equivalent to the odd (66), but where (66) resulted in oddity because of the intervention of ignorance inferences, (90) does not. This difference will be helpful.

Consider the following utterances, constructed by equivalence on the model of (65)-(68):

(91) It's completely false that France is a monarchy and that the monarch is hairy.

(92) It's completely false that I am Germanophobic and that my colleagues (do not) know it.

(93) It's completely false that I am Germanophobic and that my colleagues are Germanophobic too.

(94) !It's completely false that I am Germanophobic and that John is (not) a boche.

Be it in the current context or in an arbitrary one, a speaker uttering (94) would still be seen as displaying her negative attitudes towards German people. That is inexplicable for a presuppositional view of STs, as shown by the felicity and neutrality of (92)-(93). (94) cannot be presuppositional, and one cannot draw any sort of racist expressive content from its descriptive layer either.

Presuppositional views of STs *à la* Schlenker thus have nowhere to locate the source of that anti-German content leaking out of (94). Again, (94) is pragmatically odd for the same reasons as (85b), and again, this goes in favor of the present argument, not against it. Presuppositional accounts of STs make the prediction that (94) is neutral and non-offensive/expressive; we simply observe it isn't.

[27] By De Morgan and classical rules for double negation, $(\neg A \vee B)$ is equivalent to $\neg(A \wedge \neg B)$.

2.5. Concluding Remarks

What can we conclude from the observation that the expressive content associated with STs is harder to filter or cancel than the presuppositions of most (if not all) kinds of presuppositions?

First of all, if STs do not convey their racist expressive content via a presuppositional mechanism, this paves the way to Pott's use-conditional analysis of expressivity (Potts 2007), or maybe to more strictly pragmatic accounts (Nunberg 2016). I will turn to a consideration of these theoretical options in the following chapter.

Second, the above data could constitute evidence that the class of projective content is broader, and hence non-reducible, to that of presuppositions. That could contribute to falsify different attempts at reducing projective content in general to the narrower class of presuppositional content (see Roberts et al. 2010 or Tonhauser et al. 2013 for discussion).

Finally, we could take these results to simply indicate that, although expressive content is presuppositional, it is simply something about accommodation that makes it hard to filter. First, Heim (1983) or Van der Sandt (1992) have noted a general preference for global accommodation. Second, we saw that hard triggers like "too" are particularly difficult to filter. Third, we know at least since Geurts (1996) that the presuppositions of embedded triggers may be pragmatically inherited by the matrix sentence even when they are semantically filtered out (that is the so-called "proviso problem").

Taking these three observations together, could it just be in the end that STs (and maybe other expressives) are super-hard triggers with a really strong preference for global accommodation?

Perhaps that would not be an outlandish assumption. Just note that if STs do fit in a presuppositional picture, there appears to be pressures to globally accommodate the presupposition even when there are competing pressures to locally accommodate. That would still distinguish expressivity from standard presuppositions, and would still call for an explanation. I rather conclude, from the case study of STs, that expressive content is not presuppositional.

Facing the limitations of the initially appealing presuppositional account of STs, I will now consider some other possible hybrid accounts, before arguing that we need to consider the psychological (rather than merely linguistic) level of analysis.

Chapter 3. Two Other Hybrid Accounts of Slurring Terms

Facing the limits of presuppositional accounts of STs, I will now focus on two other families of hybrid linguistic accounts of STs. I will first briefly present Potts and others' account of expressivity in terms of conventional implicatures (CI), and show that it is better equipped to account for the projection facts, although it might import an unnecessary theoretical commitment to a reduction of expressivity to communicational phenomena, thus unjustly excluding the possibility of a mental correlate of expressivity.

Then, I will present and reformulate Nunberg's account of STs in terms of conversational implicatures and affiliatory speech-acts. I will investigate a reconstructed version of his account, and show that although it is promising, it suffers from the same flaw as other purely linguistic accounts: it closes the door to the possibility of "expressive" thought.

> One may say that there is no such thing as the proposition of belief expressed by "Nietzsche was a kraut", there is only the attitude-complex involving (a) the pure belief that Nietzsche was German and (b) a cognitive-affective attitude toward Germans. (Saka 2007, p. 143)

Within the equipment available in the linguists' tool-kit that is able to capture secondary dimensions of meaning - which could be useful to model STs -, there is another sort of conventionalized pragmatic mechanism called conventional implicature (CI). The notion of implicature is due to Grice, whose work is generally understood as clarifying the semantic/pragmatic divide (Grice, 1975).

The information that can be judged true or false by speakers corresponds, according to him, to *what is said*. Grice was interested in everything that could be communicated by utterances, and he remarked that what is said is just a subclass of all the possible information one can extract from various utterances. Some of the information that is not encoded in the truth-conditions corresponds to the implicatures of the utterance.

Implicatures come in two classes. Conversational implicatures rely on general pragmatic reasoning, or on general maxims of conversation. For example uttering (95):

(95) I will invite John or Mary.

most of the time implicates that the speaker will not invite both John and Mary. This information is conveyed by the utterance because, roughly, hearers suppose that if the speaker wanted to invite both Mary and John, she wouldn't have uttered an under-informative sentence like (95), unless she is uncooperative. These conversational implicatures are the subject of a lot of work in formal pragmatics, in order to account for each step of the process extracting this information.

On the other hand, conventional implicatures don't appeal to Gricean maxims governing conversation. For example, words like "but" and "even" are said to conventionally enrich the meaning of the sentence (Bach, 1999). The connector "but" conventionally suggests something like a contrast between the two conjuncts (see Ducrot and Vogt, 1979 for details) and the expression "even" suggests unexpectedness. Under this thesis, (96a) is equivalent to (96b) and (97a) to (97b), as far as truth-conditions are concerned:

(96) a. Mary is poor but honest.

b. Mary is poor and honest.

(97) a. Even Paul solved the problem.

b. Paul solved the problem.

CIs are often understood as comments that a speaker of an utterance makes about the content of the utterance itself. It seems perfectly in line with what we said informally about content of utterances which was somehow "whispered aside". In other words, CIs are speaker-oriented comments upon the at-issue core of an utterance.

Potts (2003) proposes the following set of properties instantiating Grice's informal idea:

(i) CIs are part of the conventional (lexical) meaning of words;

(ii) CIs are commitments, and thus give rise to entailments;

(iii) these commitments are made by the speaker of the utterance in virtue of the meaning of the words she chooses ("speaker-orientedness"); and

(iv) CIs are logically and compositionally independent of what is "said" (Potts, 2003b).

Among speaker-oriented constructions whose falsity does not impact the content that is "said", we find relative clauses ("Napoleon, whom I dislike, won the battle"), appositives ("Napoleon, the French emperor, won the battle") and pure expressive items ("That damn Napoleon won the battle"). There is in this notion of conventional implicature an idea of a multidimensionality of meaning, where communicated content comes in different strata, and this idea seems to apply well to STs.

Potts uses the idea of multidimensionality of meaning to propose a compositional analysis of the expressive level, that he identifies with the CI level (Potts, 2003a). *What is said* corresponds to standard descriptive, truth-conditional compositional principles, and *what is conventionally implicated*, including expressives and STs, corresponds to a second layer of propositional content with its own compositional rules (Potts, 2003a).

According to this view, "Obama is a nigger" conventionally implicates contempt for black people on the part of the speaker, just like (96) carries the CI that Mary's honesty is unexpected given Mary's poverty. It would then be part of the meaning of a ST that its target is contemptible, but it wouldn't contribute to its descriptive content. McCready puts it this way:

> The only way to model mixed content would be to assume that content can be introduced in two distinct stages. This idea can be implemented by assuming that pejoratives introduce an at-issue object, which is then predicated in some way by a CI object. (McCready 2010, pp. 15-16)

A CI account of STs has the clear advantage of accounting for their main linguistic feature, i. e. that they scope out, because the content of a CI doesn't belong to the content of the utterance - it is a comment about the utterance - and is therefore not embeddable, just like supplements. The expressive content of STs would be predicted to project as widely as supplements - that is wider than presuppositions-, which seems to be on the right track given the results of the previous chapter.

Potts' analysis offers a nice explanation of the projection behavior of STs by modeling two separate (although interacting) dimensions of composition. Therefore, expressives are not really embedded under the truth-conditional operators they appear under, because the CI dimension to which they belong doesn't operate at the at-issue level.

With STs, which integrate a CI/expressive and a descriptive dimension, it is therefore only the descriptive part corresponding to the neutral counterpart that is interpreted under the operator's scope.

There is an alternative CI view of STs where the stereotypes commonly associated with targets by racists play an important role[28]. Dummett suggests that we can understand the

[28] Several authors, especially Jeshion in her extensive 2013 paper about this point (Jeshion, 2013b), argue that stereotypes should not play any explanatory role. First, Richard remarks

meaning of STs in virtue of rules of inference that competent speakers master (Dummett, 1973). For instance, the ST "boche" would have a meaning constituted by the following rules of inference (approximately):

(98) BOCHE-INTRODUCTION: x is a German → x is a Boche

BOCHE-ELIMINATION: x is a Boche → x is cruel

In (98), "cruel" stands for the set of stereotypes commonly associated by Germanophobes with the target. Therefore, a speaker who is competent with the term "boche" is a speaker who is able to connect "German" and "cruel".

Consequently, on Dummett's inferentialist account, the meaning of the ST "boche" combines the denotational property of being German (in virtue of Boche-Introduction), and the non-denotational property of being cruel.

Williamson remarks that Dummett's account makes the wrong prediction that everyone, in order to master the term, has to reason according to (98). But a non-Germanophobe competent speaker does understand and master the term correctly, even though she is not willing to respect these rules, that is, to connect the property of being German with the property of being cruel[29].

that taking the stereotype account too seriously makes the counterintuitive prediction that ignoring a stereotype is linguistic ignorance. Hearers and users need not know stereotypes to understand that an utterance is offensive.

For instance, Jeshion notes that no stereotypes are associated with STs like "honky" or "goy", and yet everyone understands their offensiveness. Moreover, many stereotypes associated with a community are often neutral, or even positive. Chinese people are taken by racists to be good at math, Jewish people to be good at money management, black people to be good at sports, and so on. I do not see why only negative stereotypes would become part of the content of STs. Nunberg makes the clear-sighted hypothesis that stereotypes only intend to *legitimate* racism, which would in fact be about *alterity* itself.

[29] I will come back to this problem later, because it involves a distinction that is crucial for STs, between *mastering a term* and *possessing a concept*.

Williamson therefore puts forward a CI account as a substitute for Dummett's inferentialist account: STs are coextensive with their NCs, they conventionally implicate negative stereotypes about their targets, and such a general implicature is false (Williamson, 2009). For example, "John is a Boche" carries the false conventional implicature that Germans are, e. g. cruel - and not that the speaker believes so.

To sum up, this analysis of STs in terms of conventional implicatures could be spliced into the four following subclaims:

> (i) The offensive capacity of STs comes from their implicated component.
>
> (ii) Since this content doesn't belong to the truth-conditional dimension, it cannot fall under the scope of predicates and operators; hence it "scopes out", it "projects".
>
> (iii) The speakers' reluctance to evaluate slurring sentences as true has the same explanation as speakers' reluctance to evaluate sentences with an ordinary and false CI, whatever this explanation consists in.
>
> (iv) CIs are by definition detachable: whatever the CI term (like "but") implies is not implied by its non-CI counterpart (like "and"), and this feature predicts the existence of a NC to STs. Note that the CI thesis also has the advantage of easily generalizing to adjectival uses of STs.

I will not expand much on the CI view of STs, because I do not have much to oppose to it. It is likely to make the right predictions concerning projection, and the bulk of the argument I would like to make to move away from it relies on theoretical considerations about concepts. I will come back to this point later.

Apart from this larger theoretical objection which I will discuss later, I present the three following potential limitations of the CI account.

One potential issue with CI accounts of STs was noted by Saka (2007), who remarked that we can make an offensive and pejorative statement without having the right contemptuous attitude. But the attitude is supposed to be conventional:

> The convention would be that you only put beliefs about Germans using that overtone (or the derisory word) if you have the contemptuous attitude (Blackburn 1984, p. 149)

Someone could for instance, on some occasion, use the word "boche" without having the germanophobic attitude at the moment she is using it. This goes against the idea that the attitude is conventionally encoded in the term. This might count as a remark against the very notion of an implicature that would be conventional. Bach has indeed argued that the notion of a conventional implicature was odd in itself:

> To the extent that putative conventional implicatures really are implicatures, they are not conventional, and to the extent that they are conventional they are not implicatures. (Bach 1999, p. 338)

The objection implicit behind this remark is that, roughly, if something is conventional, the attitude is superfluous[30].

Another potential issue with the CI account of STs can be exhibited with the following contrast. Consider the following answers to A's statement involving ordinary CIs in (99) and (100), and then compare them with answers to statements involving STs in (101) and (102):

(99) A: Mary is rich but honest.

B: Yes, but there is nothing about being rich that favors dishonesty.

(100) A: Even Paul arrived on time.

B: Yes, but Paul always arrives on time.

B's responses to A in (99) and (100) might not be perfectly felicitous, but they are acceptable. In any case, it is way more acceptable than with an ST instead of a CI. With STs, the CI account would lead to exchanges like the following, between a racist/anti-Semitic speaker and a non-racist/non-anti-Semitic speaker:

[30] This objection is not so strong. The mere fact that there are derogatory utterances of STs without the speaker having the right attitude at the moment of the utterance does not make the attitude superfluous to the meaning of STs. For the first thing, we saw that there are non-weapon uses of STs. And second, even everyday literal meaning is subject to this effect: sometimes we utter things without believing it for instance, contradicting Grice's maxim of Quality.

(101) A: !John is a kike.

B: ??Yes, but there is nothing wrong about being Jewish!

(102) A: !Asia is mostly inhabited by chinks.

B: ??Yes, but Chinese people are in no way despicable for being Chinese!

This contrast between (99)-(100) and (101)-(102) is a problem for the theory of STs as CIs, because if CI belongs to a separate dimension of meaning, B's answers should be straightforwardly felicitous also in the case of STs.

Finally, although a CI view of STs might account for the main data and interesting properties of STs, an application of Grice's razor could lead us to prefer another, more linguistically economic, alternative. Grice's razor is

> a principle of parsimony that states a preference for linguistic explanations in terms of conversational implicature, over explanations in terms of semantic context dependence. (Hazlett, 2007)

In the present case, as we shall now see, it appears that most properties can be derived in terms of a purely conversational account. Nunberg has proposed such an analysis, that we will explore and enrich.

I will present a more principled objection against all sorts of purely linguistic accounts of STs later on. For the time being, let us consider another such view, relying on the notion of *conversational* implicature to explain the apparent additional bit in the meaning of STs. I will rely on Nunberg's work a lot, and introduce it with a short discussion of speech-act accounts.

3.2. Speech-act/Conversational Implicature Accounts

> We need an account of slurs as a class that explains both why they systematically perpetuate grievous harm and also how some of their core users can be ignorant of this fact; an account which writes strong negative affect directly into their conventional meaning across the board fails to do this. (Camp 2013, p. 339)

3.2.1. Speech-Act Accounts of Slurring Terms

In order to explain the expressivity of STs, we can focus not on what they say, but on what they do. Austin introduced a famous distinction between content and force, that is, between the information conveyed by a proposition, and the act that is performed by the utterance (Austin, 1975).

For instance, (103a), (103b) and (103c) express the same proposition but with a different force, they are used to perform different types of illocutionary acts:

(103) a. Go to the movies! (force of an order)

b. You will go to the movies. (force of an assertion)

c. Are you going to the movies? (force of a question)

Similarly, someone uttering "fucking American!" (2nd person use) or saying "There was a fucking American next to me in the metro" (3rd person use) performs an act of expressing a particular attitude towards Americans.

In general, as Austin insisted again and again, performing a speech-act (like "Bravo!") is not to be equated with communicating a piece of information. Such an act is purely expressive, it is not *truth-apt*. The question of truth and falsity doesn't arise because an act is not a proposition[31].

This suggests a third possible hybrid account, according to which the expressive intensifier added to a neutral term signals the performance (by the speaker) of the speech act of expressing an attitude of contempt toward the denotation of the neutral term.

In "integrated" STs (like the French "Amerloque"), a single expression would both have denotational value (like the NC "Américain") and would at the same time signal the performance of the derogatory speech act[32].

In analytical dissociated STs (like the French "Putain d'Américain"), one of the expressions would carry the denotational value (here "Américain"), and another would signal the performance of the derogatory speech act (here "putain").

[31] Although see Hanks (2011) or Soames (2010).

[32] Besides the generic act of expressing a derogatory attitude toward the denotation, several specific types of acts one can perform with a ST have been identified in the literature. Bianchi and Langton et al. identify three categories of slurring speech-acts (Bianchi, 2014) (Langton et al., 2012, Langton, 2012).

Assault-like speech-acts focus on the target and can be persecuting, degrading etc. *Propaganda-like speech-acts* focus on the addressees and are inciting hate, promoting racial oppression etc. *Authoritative speech-acts* are for Langton the acts of ranking the targets as inferior, legitimate discrimination, enacting a system of oppression etc. Jeshion (Jeshion, 2013a) adds a possible *identifying speech-act*, which amounts to taking the neutral property of the target (e. g. being Jewish) as a defining feature of the target's identity. Anderson and Lepore focus on the act of breaking social prohibitions, just like for dirty words (Anderson and Lepore, 2013).

As one shall see below, Miscevic (Miscevic, 2011), and Nunberg (citing Harris, forthcoming) (Nunberg, 2013) note the importance of the *affiliatory speech-act*, consisting in showing a complicity with a community of racist individuals. Such a wide variety of acts that STs can be used to perform supports Jeshion's point that STs are offensive for numerous reasons.

Speech-act accounts of STs have the clear advantage of explaining most, if not all, distinctive linguistic features of STs. First, the derogatory force of these devices would be directly reduced to the derogatory force of the various acts that are performed with them.

Second, since speech-acts are not propositional constituents and can thus not be embedded, the projection of the evaluative content of STs then becomes understandable. As Geach argued long ago (Geach, 1965), force does not embed.

Speech-act accounts have the other advantage over the other hybrid theories of STs that it seems to be the only account which takes seriously the "expressive" character of the so-called expressive dimension. It does not reduce expressivity to some sort of propositional, or informational, content.

In this chapter, I will focus on Nunberg's version of a speech-act account of STs (Nunberg 2013, 2017), relying on a conversational implicature. In a nutshell, according to Nunberg, STs have the powers they have in virtue of an affiliatory speech act they are used to perform. In particular, when a speaker uses a ST, she performs an act of affiliating herself with the group of (racist) individuals to whom the term belongs.

I will first provide a neo-Gricean reanalysis of Nunberg's account. After raising the problem of STs without counterparts, I will develop a novel, three-dimensional, Nunberg-like account. This will help us to better understand what is at stake with social or speech-act theories of STs, eventually leading us away from such accounts.

3.2.2. Nunberg's View

According to Nunberg, "slurs aren't special" (Nunberg, 2013). That the force of STs starts with the context of utterance and not from their encoded meaning is roughly what Nunberg argues (Nunberg 2017).

It is a misconception to think that racists use STs because - in virtue of their lexical meaning - they are derogative. In fact, STs have a derogative force only because they are precisely the words that racists use. The order of explanation is reversed.

What is offensive about STs thus doesn't come from their lexicalized meaning, but from *metadata*, that is, from encyclopedic knowledge about the term's origins, the people who use it, in which part of the world, and so on. For Nunberg,

> Slurs [...] derive their significance and force from the attitudes we associate with the people who use them. (Nunberg 2017, p. 38)

More precisely, Nunberg aims at fully accounting for the derogatory force of STs by pointing out that they are pieces of jargons. Under this view, the term "kike" has an ordinary descriptive semantics - it refers to Jewish people -, but *qua* piece of a jargon, it *belongs* to a subgroup of (anti-Semitic) speakers. "Kike" belongs to the anti-Semitic in the same way than, say, the jargon "quercus" (for oaks) belongs to the botanists.

How can the use of a jargon word be derogatory? Nunberg answers that in virtue of the principle of manner governing conversation, or Levinson M-principle (Grice 1975, Levinson 2000), the choice of a jargon word is opposed to that of a default word, and thus generates an implicature to the effect that the speaker intends to affiliates with those to whom the jargon word belongs.

For instance, the term "Jewish" being the default term to refer to Jewish people, utterances of "kike" generate an implicature to the effect that the speaker affiliates with the sub-group of speakers who own the term (the anti-Semitic), hence shares their attitudes, opinions, dispositions towards Jewish people[33].

[33] Note that, what it is for a term to *belong* to a linguistic community, what a linguistic community exactly consists in, or what it is for a term to count as a *default* term, are subordinate questions that Nunberg does not address directly.

I will admit for the sake of discussion that there are independent, non-circular criteria that one can rely on to provide an account of these more basic phenomena. I will just remark that the relevant notion of *linguistic community* must be intentionalized for it to paly the role Nunberg needs it to play, for members of a linguistic community might use both their term and the default term depending on the occasion. A term *belonging* to a community might be identified with the term being used exclusively by its members, but must be distinguished from its members using it exclusively.

The view thus treats the contrast in (104) and the one in (105) along the same lines:

(104) a. Three Germans are walking.

b. !Three boches are walking.

(105) a. Three robins are flying.

b. Three *erithacus rubecula* are flying.

According to Nunberg's account, (104a) and (104b), have the same descriptive meaning, just like (105a) and (105b): they are true in the same situations, false in the same situations. "Erithacus rubecula" is the scientific term for robins - it refers to robins, and "Boche" is the germanophobic word for Germans - it refers to Germans.

The contrast between (104a) and (104b), that is, the additional effects that the use of "boches" trigger in (104b), have the same source as the effects that the use of the scientific jargon "Erithacus rubecula" in (105b) trigger. Where (104b) usually conveys derogation or offense, (105b) may convey pedantry or condescension, and they do so in virtue of one and the same mechanism: inferences drawn after the violation of the maxim of manner.

In both cases, the vocabulary that is chosen in order to make the reference job is "deviant" vocabulary. When there exists a default way to make reference to something, it brings surprise to borrow a term from a scientific dialect which does the same descriptive job (at least in the context of everyday ordinary talk).

I may do so if I have for example, in addition to my *primary intention* to communicate that p, a *secondary intention* to show how erudite I am, or maybe to mock a scientific approach to everyday things, or something along these lines[34].

In any case, there must be a *reason*, over and above my intention to communicate that p, why I do so in a non-standard manner. Nunberg suggests that it is the pragmatic system, aimed among other things at figuring out the reasons speakers have to deceive expectations,

[34] I will elaborate on the distinction between primary and secondary intentions later in the present section.

which is responsible for the derivation of the powers of STs to offend through a (affiliatory) speech-act.

This, Nunberg says, is enough to accommodate all features of STs, from their projection profile to their effects in communication. As he puts it,

> ... slurs tend to be speaker-oriented because they are marked alternatives to a conversational default, so the speaker always has an ulterior reason for using them, over and above the proposition he asserts. (Nunberg 2017, p. 37)

And that reason usually is the speaker's intention to affiliate with a group. Let us now focus on Nunberg's account and reformulate it in a neo-Gricean manner, to consider how it is armed to address the main issues that STs seem to raise. I start with a discussion of the issue of cancellability.

3.3. Toward a Neo-Gricean Nunbergian Account of Slurring Terms

3.3.1. Cancellability

An immediate question arises: if the effects of STs are primarily the product of conversational mechanisms - and since conversational implicatures are typically cancellable (Grice, 1975), why are they hardly, if ever, cancelable?

Take an utterance of (106), which implicates that Mary got a promotion after she bought an apartment, even though its literal meaning (provided a classical analysis of conjunction) does not specify an order:

(106) Mary bought an apartment and got a promotion.

When a qualification is added to (106), like the one in (107), the implicature is cancelled:

(107) Mary bought an apartment and got a promotion, not necessarily in that order.

Now compare the above pattern to the one below, featuring STs in a similar environment. Whatever content is conveyed by the use of a ST, it seems hardly cancellable by any sort of qualification:

(108) !?Three kikes are walking; note that I'm not Anti-Semitic.

(109) !?There is a boche downstairs; and everybody knows that Germans are adorable individuals.

(110) !?I'm sitting next to a chink in class, although you and I are not prejudiced against Chinese people.

(111) !?My sister dates a nigger, and I'm perfectly fine with that as African-Americans are just like everyone else.

If the effects of STs really are the result of an implicature, as Nunberg claims, why do (108)-(111) still convey disparagement towards their targets? Why aren't these implicatures cancellable?

In fact, there are independent reasons to expect manner implicatures to be hard to cancel. Grice even recognized the specificity of manner implicatures in this respect (Grice, 1989).

For example, the use of the complex expression "caused to die" instead of the standard and simple "killed" usually implicates the absence of intention to kill. Now, it is very hard to imagine a context in which the most salient reason the speaker has for using "caused to die" instead of "killed" is something other than his intention to communicate the absence of intention to kill:

(112) ?John caused Mary to die, and I'm sure he did it intentionally.

It might be slightly improved in (113), if uttered by the judge of a trial for instance:

(113) Waiting for evidence establishing whether he did it intentionally or not, the court acknowledges for the time being that, at least, Mr. Smith caused Mr. Clark to die.

We see here that the context must meet constraints that are so specific to allow the implicature not to arise that it is expected to be very robust, as opposed to the implicature of (106) which was easily cancelled in (107).

But even very robust, manner implicatures are in principle still cancellable, because they follow from the incorporation of pragmatic contingencies into the calculus of the (enriched) meaning of an utterance.

Now, we might even find cases where the conditions for an inference to be cancelled are so improbable or hard to meet that it could seem to the imprudent observer that the implicature is not cancellable. It is nevertheless important to distinguish between cancellability as an abstract property (that all implicatures do meet), and cancellability as instantiated property (that some implicatures might not meet)[35].

[35] To make the point clearer, the distinction is analogous to the one in the philosophy of science between a hypothesis being *falsifiable* in principle, and its being actually falsifiable with today's technological means. In principle, a theory might be falsifiable in the sense that

In principle, since the mechanism that Nunberg calls on relies on hearers' attempt to figure out a reason why the speaker said things in a deviant manner, nothing prevents the specific implicature to be cancelled if another reason were a better candidate for explaining the speaker's verbal behavior.

The implicature derived from the use of "caused to die" instead of "killed", or the implicature derived from the use of "boche" instead of "German", might indeed be extremely robust, but in the technical sense, they are still cancellable.

Understanding why and how different manner implicatures come to be hardly or never cancellable is an interesting project *per se*, but without having to pursue it, Nunberg dismisses the potential objection of the non-cancellability of STs' effects in acknowledging that it is just a general property of manner implicatures.

A case where STs might be used without an affiliatory intention are echoic uses, like "this kike is going to kick your ass" as an answer to an anti-Semitic ST, or "This guy changed seats because he did not want to sit next to a kike". Note that the fact that these uses are echoic does not disqualify them. These examples are perspectival and would require a lengthier discussion to see in what sense they constitute cases of cancellation. I discuss such issues in the appendix to chapter 7.

Taking stock, we just saw that, just like manner implicatures, the social effects of STs are hardly cancellable, although they are cancellable in principle. Nunberg's account of STs seems to be on the right track so far. Indeed, why else than in order to impersonate a member of a group g would a speaker use a non-default term t that belongs to g? What reason other than affiliation with the anti-Semitic could there possibly be for using of a term borrowed to the anti-Semitic? Other reasons are rare, hence the (false) appearance of non-cancellability.

Note also that manner implicatures, unlike most types of implicatures, are not nondetachable. Nondetachability is the name of the impossibility

> to find another way of saying the same thing, which simply lacks the implicature in question. (Grice, 1975).

---

one can conceive of an experiment that would falsify it, even if that experiment were impracticable.

As manner implicatures are triggered, precisely by a property of the "way of saying", saying the same thing in another way typically destroys the implicature. This is still in accordance with what one observes in the case of STs, provided that "John is a Boche" is "another way of saying" that John is a German.

3.3.2. A Nunberg-Like Account

Nunberg's account relies on conversational, Gricean, principles. In a neo-Gricean manner, we can thus attempt at giving it a slightly more formal character. Doing so will help us identify specific predictions of the view and discuss more precisely its potential shortcomings. Let us first introduce the default maxim:

Default Maxim: Use default vocabulary unless there is a reason not to.

This maxim supposes that being default is a feature that a term might have or not have. Note that the notion of default is a contrastive notion. If there exists in a language only one term t to refer to x, then t cannot be default. Only when two terms refer to x can one of the two acquire the status of being default[36].

That a term t1 is the default makes it preferable to the coreferent alternative t2, so that speakers expect t1 to be used in order to refer to x, rather than t2. Nunberg suggests that STs are just alternatives to such a contextual default.

On this basis, for the effects of STs in conversation to be derived, some elements must thus be in place. First of all, there must exist in the language two alternative terms[37]:

[36] Nunberg alludes to mechanisms of social or political negotiations, by which a term might acquire the status of a default. The diversity and specificities of these mechanisms are not essential to the present discussion, as what matters is that one of the terms available does have the relevant conventional default status.

[37] We don't consider cases where there are more than two terms referring to x, for the sake of simplicity.

- Alternative terms: {t_1, t_2}

and one of the two must have obtained the status of being default (relation represented here as an ordered pair, with t2 being the default):

- Establishment of a default: <t1, t2>

These, plus knowledge of the provenance of t2 (that is, the linguistic community who owns the term), constitute the prior knowledge that Nunberg's view assumes is necessary for speakers to derive the relevant effects of STs through an affiliatory speech act. In virtue of the default maxim, an utterance U2 that contains a default term is preferable to an utterance U1 involving a non-default term.

We mark that relation U2 >> U1. Because t1 and t2 are alternatives to each other, any utterance containing one or the other generates an alternative, just like in the case of scalar implicatures:

- Utterance U: "... t1 …"

- Alternative A: "... t2 …"

When the default maxim is applied, we obtain A >> U. Nunberg remarks that in the case of STs, a speaker utters a sentence that has a preferred alternative because of her intention to affiliate with the sub-group to whom the alternative term belongs, that is, an affiliatory speech-act is derived: intention of affiliation is usually the reason why a non-default term is chosen. Interpreters must therefore have knowledge of the term's provenance in order to be able to make the right inference.

Overall, we obtain the following computation for an utterance of, say, 'John is not a kike', by a speaker s (steps in between brackets below represent to the prior knowledge that is needed for the Nunberg-like computation to be successfully performed):

[Alternative terms: {kike, Jewish}]

[Preference relation: <kike, Jewish>]

[Provenance of the non-default 'kike': the anti-Semitic]

- Utterance U: "John is not a kike"

- Generation of an alternative utterance U': "John is not Jewish"

- Application of the Default Maxim: U' >> U

- General reasoning: s intends to affiliate with the anti-Semitic (affiliatory speech-act)

The above example involves a negation. We can thus see how the right projection pattern can be derived: since the pragmatic mechanism is, under this view, triggered by the use of the term (rather than by its meaning), together with some prior knowledge about the term's alternatives and its provenance, all linguistic environments in which the term is really used[38] will end up having the same effects in conversation.

Now, as we shall see below, not all STs have a default alternative, and it seems to be possible to affiliate with a group using its vocabulary without there being an existing default, provided that piece of vocabulary is identified as belonging to the group. What role then does the default maxim play? Why does Nunberg place a default constraint on the interpretation of the offensiveness of STs?

By introducing three clarifications and distinctions, I will now try to improve and make explicit the essence of Nunberg's view, proposing that the presence of a default is necessary for the intention to affiliate to be interpreted as a *primary intention*, as opposed to a secondary intention.

[38] And not necessarily when the term is mentioned, or echoed. See Bianchi 2014 or Recanati 2007 for more on this aspect.

3.4. Three Clarifications

I provide below three sets of clarifications. The first distinguishes between two levels of communicative intentions. The second stresses the importance of distinguishing the users of STs from the rest of us in the pursuit of their meaning. The third fleshes out three levels of inferences that stay somewhat undistinguished in Nunberg's account.

3.4.1. Primary Intentions and Secondary Intentions

> If pejoratives do indeed carry colouring conventionally, it is partly because they exist in the language as alternatives to other words with the same denotations. Why would a speaker call a person a "faggot" rather than a homosexual, or a "nigger" rather than a Black or African-American? This choice of terminology is explained by the intention to express contempt towards a group. (Finlay, 2005, p. 13)

A distinction between primary communicative intentions and secondary communicative intentions is helpful here. Say I am taking an umbrella to go out because I do not want to get rained on. Now as it happens, taking the umbrella will also prevent my roommate from taking it if she wants to go out too. Of course I do not really intend that second consequence, even though I may be aware of it.

My primary intention was to not get rained on, and since that could not go without preventing my roommate from taking the umbrella, this later intention was just a secondary intention. I do "want", in a certain sense, to prevent my roommate from taking the umbrella, as I am morally responsible for that consequence, but it is not that I prefer that this happens rather than not.

On the contrary, I do prefer worlds in which I do not get rained on to worlds in which I get rained on. An ideal world might be one in which I do not get rained on and my roommate can take the umbrella too, but in that case, it just so happens that such an outcome is out of reach.

Now, let us suppose that there was an action at my disposal that was not more costly than taking the umbrella and could avoid the bad consequence. Say for instance that I could have taken my old raincoat instead of the umbrella (ignore the fact that in practice an umbrella might be more useful or pleasant than a raincoat etc.). Only then, if I stick to the umbrella could one conclude that I might in fact want my roommate to get rained on somehow. This is thus no more a secondary intention, and becomes a primary intention.

Importantly, note that it can be interpreted as a primary intention only because there was an alternative at my disposal. Applying the same reasoning to STs, we see that, even though we can affiliate to a group by using its vocabulary without there being a default alternative, the existence in the language of a default alternative is necessary for that intention to be seen as primary.

And it might well be that only a primary intentions to affiliate with a racist group constitutes a slurring speech-act, which would enlighten Nunberg's insistence on the defaults.

3.4.2. Meaning for Users, Meaning for Others

There are (at least) two (related) dimensions about the meaning of particular terms. Bits of meaning, in the broad sense, depend on external factors and are not necessarily cognitively represented by users; others are internal and occur at the personal (or sub-personal) level.

Under internalist approaches to meaning (e. g. Fodor, 1981, 1982, among many others), the meaning of a term t consists solely in (parts of) whatever is represented by its relevant users. Users of a term thus play a theoretically crucial role in an investigation of the term's meaning properties.

STs are the words that the racists use, and that we do not use: it looks just like they are part of a dialect. Now of course, it's not as if STs were part of another language that we do not understand at all: even though we do not use them ourselves, we do have some competence with them. We understand sentences involving them, are aware of their offensiveness, their (intended) reference and so on.

When linguists and philosophers investigate the properties of a language system, they focus on the judgments of acceptability, truth or felicity of individual speakers of that language. This should not be different for STs. As STs belong to a subgroup, an investigation of their meaning properties should primarily focus on how the speakers who use the term encode and represent its different components.

Understanding the precise nature of the competence that out-groups display with an in-group term is a different project. Indeed, there are some words of other dialects whose meaning I do not know at all, others for which I have some clues how they function, and others that I don't use myself but that I master perfectly. To what extent do most speakers master slurring terms?

If, as Nunberg's account suggests, the offensiveness of STs has to do with the deliberate choice of a deviant lexical item departing from a negotiated default, it then follows that STs can be mastered and interpreted only by speakers who belong to, or at least understand, two dialects: the racist in-group dialect, and the default out-group dialect.

It is indeed only when one knows that there is a standard term that one can infer that the use of a slurring term purposefully departs from it, access to the derogatory content, and thereby that we can mean it to be derogatory. A direct prediction of Nunberg's account is thus that a term cannot be derogatory unless speakers have access to an alternative, co-referential term.

As we just saw and will discuss more below, that prediction is problematic given the different cases of STs either without a known lexicalized alternative default, or of STs used by racist individuals who do not know that there exists an out-group alternative term.

In any case, two dialectal communities must be kept apart in accounting for the behavior of STs. There is a *supra-community* of speakers mastering the linguistic or social conventions of negotiated-English to call members of a group g "X" (the out-group), and an *infra-community* of racists speakers for whom it is conventional to call the target with another term "Y" for members of g (the in-groups).

"X" and "Y" are co-referential, but the respective groups which the convention ruling their use is relevant for differs. Offensiveness could very well come in some cases from the fact that speakers who slur in general have a reason to prefer "Y" to "X" for gs: for example a refusal of the negotiated default "X", or an affiliation with the sub-community of "Y"-users.

But if the phenomenon of STs had to do only with this confrontation between two dialects, a term would be a slurring term only in cases where there is such linguistic war. Before such a confrontation between two terms belonging to different groups were to take place, there could be no STs at all, even in clearly racist groups despising and actively discriminating a target.

Again, for Nunberg, a term "Y" is a ST only if it is used as opposed to an alternative "X". It is just conventional for a group G1 to call members of g "X", and conventional for another (sub-)group G2 to call members of g "Y". Since members of G2 share contempt for members of g, preferring "Y" to "X" constitutes an offensive affiliatory speech-act.

In a nutshell, the present objection is that we need to distinguish between two relevant linguistic communities: the infra-dialectal community, where the term is a ST already, and the supra-dialectal community, where the lexical fight takes place. The intergroup lexical war cannot precede nor explain the offensiveness STs, if STs pre-exist at the subgroup level.

3.4.3. Natural Meaning, Affiliation and De-Affiliation

Consider the following:

i) a term t belongs to a group g;

ii) members of non-g have a default term d.

iii) it is common knowledge that i) and/or ii)

i), ii), and iii) are distinct facts that can be associated with inferences of different nature. A slurer could be aware of all of them, or only of two of them, or only one, or even none of these facts. Depending on properties of the context of utterance, hearers might come to ascribe any of these states to the speaker.

By conflating these three levels, Nunberg focuses merely on cases where a term has effects in virtue of being uttered by a speaker mastering i), ii) and iii). But the other states are logically possible, do exist in natural situations, and do trigger some peculiar effects in communication that it is relevant to look at when focusing on STs.

Faced with the use of a ST, we can indeed distinguish between three types of inferences about the speaker's mental states. First of all, the basic step in a deflationary account *à la* Nunberg is the recognition that a certain term that is used belongs to a community.

Based on that very basic fact, independent of whether a speaker has any intention to offend, to affiliate herself with a group or anything of that sort, the very fact that she uses such vocabulary can show that she belongs to the group. This is not strictly speaking an implicature, as implicatures are inferences about the speaker's communicative intentions, that is, intentions to make some information common knowledge.

Here, we rather face an instance of Grice's *natural meaning*: the mere usage of a word, a word which happens to belong to a certain group g, indicates that the speaker is a member of g, just like an Italian accent indicates that the speaker is Italian, or like blushing indicates that the speaker feels ashamed. The information is shown, or displayed, rather than communicated, it is not part of what the speaker said in any relevant sense, and one need not reason on the speaker's intentions to access it.

Take for instance the following real life case, keeping STs in mind. People from southwest France use the term "chocolatine" to refer to a chocolate-filled pastry that the northern half of France calls "pain au chocolat"[39]. When a customer in a Parisian bakery orders a "chocolatine" (note that Paris is in the northern half of France), hearers typically infer that the speaker has just arrived from southwest France, independent of her awareness of this linguistic difference, and hence independent of her potential communicative intentions.

Hearers can retrieve information from the use of a term, even though this piece of information is not part of what is communicated, not part of speaker's meaning. A big deal of

[39] Thanks to B. Spector for his insight on this example (p.c.).

Nunberg's "metadata" account relies on such Natural Meaning inferences. We know something about the word's provenance, so that hearing someone using it is in itself indicative of something about her.

Note that this inference widely projects out of truth-conditional operators. Similarly with STs: when a speaker uses the n-word, independent of her intention to offend or to affiliate herself with a group and so on, as the n-word belongs to the racists, her use of it shows naturally that she belongs to the group of racists. It need not be part of the speaker's knowledge that the term belongs to the racists or that there exists an alternative coreferential term.

Only on a second stage can the speaker know that there exists another term for the same pastry, the expression "pain au chocolat", and then purposefully decide to stick to "chocolatine", possibly with the intention to make her decision manifest. In this case, if the speaker is aware of the existence of "pain au chocolat", then hearers will recognize that "chocolatine" was used on purpose and recover a reason for this lexical choice.

But it might not be the case that "pain au chocolat" was recognized as the default term. Maybe at this stage, no constraint on the preference of "pain au chocolat" over "chocolatine" is inscribed in the common background.

The reasons why the speaker decided to stick to "chocolatine" might be because she intends to affiliate herself with southwest France people, independent of any intention to depart from another convention. That is, the speaker might know that there is an alternative, without knowing that this alternative ought to be locally preferred. This second sort of inference is of a different type than the first one, as it is now part of what is communicated: it is an implicature.

On a third stage, the speaker could acknowledge not only that there is an alternative expression, "pain au chocolat", but also that this alternative ought to be preferred because it has the status of a default. Then, in addition to her basic reason to affiliate herself with people from southwest France, an additional reason why a speaker would use the non-default "chocolatine" might be to de-affiliate herself with the Parisian local group, provided that it is common knowledge that she knows that "pain au chocolat" is locally preferable.

When that last stage is reached and manifest in communication, hearers can retrieve the following information from the speaker's use of "chocolatine" in a Parisian backery: i) she is from southwest France, ii) she intends to affiliate herself with people from southwest France (primary implicature resulting from the explicit choice of lexicon), and iii) she intends to de-affiliate with people from Paris (secondary implicature resulting from a violation of the default maxim). Nunberg's view focuses on the second level, and it is well equipped to account for the third level. But as we began to see, it is under-equipped to account for the first level, that is, for the offensiveness and expressivity of racist expressions without alternatives.

I show now how a Nunberg-like account can succesfully separate the three dimensions in giving less weight to the counterpart condition. The distinction of these three levels of inference, along with the above two clarifications (primary vs. secondary intentions, and the two distinct linguistic communities), leads us to formulate a modified Nunbergian account.

3.4.4. Toward a Three-Dimensional Nunbergian Account of Slurring Terms

Here is a chronological/logical Nunberg-like computation of the three levels of STs' derogatory meanings described above, taking into account the previous three clarifications and improvements. Steps between brackets correspond to prior knowledge that is necessary for speakers to compute the meaning enrichments of each level. Note also that t' might be the empty-set, that is, there might be no alternative way to refer to the target:

A: s used term t1

[It is common knowledge that t1 belongs to infra-group g1]

Level 1 (natural meaning): s is a member of g1.

[s knows that level 1 will happen, the interlocutor knows that s knows that, and so on]

B: level 1 is common knowledge.

[It is common knowledge that a term t2 has the same descriptive content, is not more costly, and is used by the supra-group g2. s could have used t2 and avoid B]

Level 2 (default maxim): that B is a primary intention, s primarily intends to de-affiliate with g2 and/or to affiliate with g1.

[s knows that level 2 will happen, the interlocutor knows that, and so on]

Level 3: it becomes common knowledge that it was s's primary intention to communicate (make common knowledge) his or her belonging to group g1.

This reformulation of Nunberg's deflationary, social proposal is aimed at furthering his attempt to reduce the expressivity and offensiveness of STs to the recognition of (primary) affiliatory intentions.

Whereas the initial account Nunberg proposed was bound to treat cases of STs without a neutral counterpart as exceptions, the mechanism I just described has the resources to account for these. It involves many dimensions: the provenance of terms, competition between lexical items, the notion of a default, primary and secondary intentions to affiliate with groups of speakers, inferences about speakers' intensions and other mental states, and so on.

Although the existence of all these dimensions has independent motivations, the engine can look quite heavily loaded for a deflationary account of STs. But the account still is deflationary in the sense that nothing in the dictionary meaning of STs indicates conventionally that it is a ST, contrary to most other hybrid theories that Nunberg gathers under the heading of "utterance-condition view":

> On the utterance-condition view, it's conventional among English-speakers to use 'nigger' to refer to blacks in order to express racist attitudes [...] On my view, roughly, it's a convention among certain English-speakers who have racist attitudes to use 'nigger' to refer to blacks (Nunberg, p. 46)

For Nunberg, we understand that the term is derogatory not from its conventional meaning, but from its metadata, that is, from knowledge we have about the word itself as an object.

We know who uses the term, we know what the expression is used to refer to, we know that the term is in competition with other terms, we know that the term is not standard, and so on and so forth. This knowledge, plus a ritual Gricean mechanism for extracting speaker's intentions seems to be able to provide the right results.

I provided above a sketch of a rephrasing of Nunberg's account in order to flesh-out its different components. I showed that at least three different conversational principles were at stake. First, participants detect departure from a default; second, they infer that the speaker intends to impersonate a member of sub-community; third, they infer that the speaker intends to de-affiliate with the preferred group. At the end of the day, Nunberg's view requires that

> ... words can only function as slurs if the language offers a non-slurring synonymous word (Nunberg, p. 29)

Indeed, in our above reconstruction, it is possible for a term to count as a racist offensive term even when there is no alternative in the language (level 1). There are two sorts of such cases. First, there exists many STs without an obvious default alternative term in the language, and second, we can conceive of a racist individual using a ST to insult her target without knowledge of the (possibility of an) alternative. I now turn to these two cases.

First, sometimes, offense lies in the very fact of possessing a term for a diversity of individuals that it does not make any sense whatsoever to classify together. In these cases, there could be no default term with the same reference. The term reveals that the speaker adopts a certain classification scheme that others do not. For instance, saying something like "She is like most orientals." reveals that the speaker belongs a a certain group, presumably, but there is no clear counterpart, and no primary intention to affiliate on the part of the speaker.

The English "dark-skinned" and the French "personne de couleur" might also be instances of exactly such a case: they apply to a diversity of individuals that it makes no demographic sense whatsoever to put together. Just like any use of a noun, like "chair", presupposes the existence of a relevant criterium gathering, say, a brown wooden and a white plastic object, any use of the term "dark-skinned" presupposes the relevance of a categorical criterium putting different individuals together.

This presupposition might be offensive in itself. In this sense, the mere possession of a term that is based on an irrelevant property might trigger offense. The French term "asiatiques", putting together individuals of many different countries, could in this sense be perceived as a ST.

Another example, from the political domain, is the term "populist", which seems to equally apply to people from the far right to the far left, as soon as they are perceived as speaking to the people's irrational passions. These terms presuppose the relevance of a certain categorical criterium, and do not require the existence of an alternative term to trigger the same sort of offense as other racist STs like the n-word.

In the above reconstruction of Nunberg's account, we would not need to go beyond stage 1 to adequately describe what happens in these cases. The contingent existence or inexistence of other linguistic communities with different attitudes towards the same individuals and with a term to refer to them, is in no intuitive way a necessary condition for the use of an expression to be racist and offensive.

The departure from a preexisting negotiated default is a possible additional factor conspiring to trigger offensive implicatures in some cases, but it need not be a defining feature of STs.

There are many other examples of STs without an obvious, lexicalized neutral counterpart. Here are some: "gook" (for Korean or Vietnamese), "yuppie" (young urban professional), "yellow cab" (Japanese women who only date non-Japanese), "Anchor baby" (American born Mexican whose parents crossed the border illegally), "Ainu" (native Japanese islanders from Hokkaido), "abc" (American born Chinese who is taken to not understand Chinese culture), "abcd" (American born Indian who is taken to not understand Indian culture), "abco" (Aboriginals who are alcoholics), "amerikos" (Russian term for Americans), "stinkpotter" (kayaker's term for persons using a motorized boat), and so on.

3.6. Three Limitations of the Neo-Nunbergian Account

3.6.1. Cocooned Communities

There is a kind of uses of STs that the account Nunberg proposes, even under the reconstructed version, under-generates. Imagine a remote, cocooned community of racist subjects sharing the same sort of contempt and attitudes towards their target. Imagine they have never met any non-racist community, say, they have slaves of a certain demographic category, it has always been so and they have never even though about the possibility of another system.

Nunberg's account predicts that in such a community, the term they use to refer to their target is not a ST. Indeed, as there is a lack of group dynamics, even if speakers of that community have several different words for their targets, none could be a ST because there is no negotiated default term.

Without an out-group community to linguistically interact with, all terms referring to their target fail to be contrasted with a potential alternative term of the out-group. Speakers can thus not use the term with the intention to affiliate themselves with any group whatsoever; Nunberg's mechanism or (primary) intention recognition is blocked.

Whether or not speakers of that community "slur" their target might not be a truly meaningful question, but at least Nunberg's deflationary approach is committed to answer "no": speakers of this remote cocooned community do not interact with other non-racist speakers and are unaware of the existence of a negotiated default; they lack the knowledge that there exists a term alternative to theirs. When they use the ST, they therefore slur independent of any act of self-affiliation with a subgroup of speakers, or rejection of the negotiated default.

But it seems that ruling out their speech act as a "slur" is arbitrary. We could as well want to include it in the explanandum. After all, as speakers of that community despise and hate their target, why couldn't they slur them? Does slurring really require an interaction with an out-group, non-racist community? Isn't racism more primitive, more visceral than that?

If it is, then Nunberg's account of STs would leave open the possibility of inexpressible racism. One could be racist towards a group g without being even able to express one's racism towards g (at leas to express it in as direct a fashion as with an expressive word). One could have a racist concept of the targets that one could not express with a single word, lacking knowledge of an alternative default. This is implausible, but hopefully there is another way to go.

I think speakers in our toy example can in fact slur, and they can, not in virtue of inter-dialectal dynamics, but because of a more primitive reason: they miscategorize their targets with misplaced non-conceptual, social and/or emotional dimensions. Nunberg might want to restrict a theory of STs to offensive terms that always have alternative available term, but I do not see why that should be a desideratum.[40] A more general account is needed.

And there are real life examples approaching our idealized scenario involving the use of an offensive slurring term without knowledge of a salient alternative. Consider for instance the case of a child coming back home and uttering the n-word. Her parents are upset and ask her not to utter that word ever again.

At that moment, the parents and the child belong to slightly distinct linguistic communities, and the child is faced with an alternative. Either she accepts the convention of her parents, or she accepts the convention of some of her comrades. She therefore has to choose the linguistic community she wants to belong to when she speaks[41].

According to Nunberg's account, only after she chooses to stick to the n-word will her utterance constitute an offensive affiliatory speech-act. But even before making that choice, if the child came back from school, not only with new vocabulary, but also with the associated racist feelings and attitudes, don't we want to maintain that her very first utterance

[40] And if Nunberg's point is just that the English American term "slur" applies only to that subclass of expressions his mechanism identifies, then it looses its theoretical interest on the nature of expressivity.

[41] I am not supposing that this is conscious activity.

of the n-word was an instance of racial slurring? When a racist utters the n-word, why would it be necessary that she is aware of other people's conventions for it to count as an instance of a ST?

3.6.2. Expressivity

Slurring words like "kike" have strong similarities with complex derogatory expressions like "dirty Jew": they seem to be equally anti-Semitic and derogatory, their pejorative force equally scopes out of negation and other truth-conditional operators, and so on, they are used by roughly the same groups and so on.

Now, if we hold the view that "kike" has the effects and force it has because that term belongs to the anti-Semitic, and at the same time wants to maintain that "kike" and "dirty Jew" are synonymous, then we are led to the view that "dirty jew" belongs to the anti-Semitic in the same sense. This is doubtful, as both the (expressive) modifier "dirty" and the predicate "Jew" are common English terms which do not belong to any particular infra-community.[42] Is the expression "dirty Jew" so different from "kike"?

Nunberg seems to restrict his use of the term "slurs" to terms whose specific effects follow from the sort of affiliatory speech-acts described above, where the evaluative import of the term isn't part of the conventional encoded meaning of the term. On the other hand, Nunberg

[42] It is logically open that it is the composition of the two common English terms that belongs to the subgroup, thus enabling the complex expression to trigger a jargon effect *à la* Nunberg. There are such cases in natural languages outside of slurs. For instance, the French determiner "d'aucuns", meaning something like "some indefinite individuals", is used only by a very narrow and specific group of people (roughly upper-class old fashioned well-read individuals). On the other hand, both the preposition "de" and the negative quantifier "aucun" are common in everyday French.

does recognize the existence of expressives like "damn" or "fucking", and of thick terms that "mix classification and attitude",[43] as he explicitly contrasts them with STs.

He indeed discusses how mere "prejudicials" (among which he puts STs) can become thick terms by encoding the attitude associated with the term. He notices that thick terms come with a sense of redundancy when followed by an explicitation, like in "Toadies are obsequious", where mere prejudicials do not give rise to such redundancy, like in "Kikes are bad/greedy/etc.". Redundancy here works as a test for encoded meaning: what is redundant is already encoded at some level in the term (which will be called "thick"), what is not is not conventional (and the term is a "prejudicial").

Nunberg also remarks that many STs, like "bitches", are ambiguous between a prejudicial reading (when targeting women in general) and a thick, expressive, reading (when targeting only a subset of women). In that case too, he uses a test of redundancy or contradiction to diagnose whether the term's import is the result of thickness or of an implicature.

Under the thick reading of "bitch", it is redundant to say that "these bitches are nasty" and sort of paradoxical to say that "these bitches are chaste", whereas on its prejudicial reading, it can be informative to utter that "these bitches are nasty" or that "these bitches are chaste".

So Nunberg's view about STs is not a general deflationary approach on the so-called expressive dimension, it seems to be mostly aimed at redirecting attention to non-conventional, social and pragmatic ways of giving rise to psychological effects that look like expressivity, pejorativeness and so on. So Nunberg acknowledges the existence of expressivity, and does not present a general view of the expressive dimension of language.

But if thick terms really are "thick", in the sense that they possess a conventional descriptive component and a conventional evaluative component, and if many prejudicials do become thick terms by a process of conventionalisation, then it seems that Nunberg's opposition to hybrid conventional accounts of STs is rather terminological.

[43] Nunberg expresses (p.c.) discomfort with the notion of a "thick term", as it is opposed to alleged "thin terms" (e. g. "good" is often seen as a thin evaluative, and "table" as a thin descriptive term) whose existence he doubts.

Nunberg's point would merely be that the American English term "slur" applies only to these terms whose import is not yet conventional. But why would we want to restrict an analysis of expressivity in natural language, and of STs, to just the words that still work through pragmatic reasoning and whose derogatory content has not been conventionalized yet?

Nunberg is right to warn us that such cases exist, that there are many, and that probably not all "slurs" should receive the same treatment. But our goal is broader: not only do we want to understand how STs and other "prejudicials" work, we also want to understand expressivity which is encoded, such as in "Bitch" in the thick sense, or in the analytical ST "dirty Jew".

Nunberg's deflationary account does not provide us with resources to understand the diversity of these cases. Could the similarities between "kike" and "dirty Jew" be a mere accident? Why would they occur basically in the same sets of (racist) contexts? Why would sentences where they appear both strike us as similarly anti-Semitic? And what about thick evaluative terms like "bitch"?

In a nutshell, the argument is: is it part of the meaning of "dirty" to turn something into a ST? If yes, we do not understand the near-synonymy between "kike" and "dirty jew", since Nunberg's mechanism does not seem to be relevant to understanding "dirty".

3.6.3. Autonymic Connotation and Conventionalization

At this stage, we should recall that the speech-act account of STs introduces an important link between social dynamics and conventions in language. STs appear to be used in order to perform loads of different acts. But this can mean two things.

On the one hand, some words conventionally mark a certain force (like "Thanks!" or "Bravo!"), and on the other hand, some utterances have a certain force, not in virtue of their words' encoded meaning, but merely because of what they contextually convey. Take for example (114):

(114) I will come tomorrow.

(114) is just an assertion, but it often has the force of a promise when uttered. Similarly (115) is merely a question, but is usually interpreted in context as having the force of a request:

(115) Could you pass the salt?

Is the force of (the affiliatory speech act triggered by) STs a contribution of the word's encoded meanings (like "Bravo!", or like the fact that (114) is at some level an assertion and (115) a question) or is it merely a consequence of its meaning relying on some additional, maybe social, mechanism (like the fact that (114) is at some upper level a promise and (115) a request)?

The mechanism Nunberg is calling on is reminiscent of an old phenomenon evoked in Rey-Debove, J. (1978) (but identified earlier) and coined "autonymic connotation". The autonymic connotation of a term consists in the inferences we can draw from the use of the term based on some prior knowledge one possesses about the customary users of the term.

The case of "chocolatine" would be a typical example of autonymic connotation, because we can draw the inference that the speaker is from southern France based on the prior knowledge we possess about the customary southern users of the term.

At a first stage, it can be merely an instance of natural meaning: the speaker shows where she comes from in virtue of her very use of the term. At a second stage, it can be an instance of non-natural meaning if the speaker intends to communicate her provenance in making manifest her deliberate choice of the term. For the autonymic connotation to be the result of an implicature (rather than from mere natural meaning), it is thus a necessary condition that the term is used as a non-standard manner.

It will be precisely because it deceives a shared expectation that the use of the term can be used to make manifest (common ground) an intention to communicate one's affiliation with a certain sub-group. The non-default character of the chosen term is therefore essential to the effects it triggers in communication.

Now, as soon as such a manifest speaker-meaning is present, then regular mechanisms of conventionalization must be at play. The term, as it is used in communication over a certain period of time, should see its connotation conventionalized, just like other implicatures get conventionalized.

But in the case of autonymic connotation, there seems to be two opposing forces at play in the process of conventionalization. Being used more and more often, the non-standard basis triggering the implicature vanishes, and two things might happen. Either autonymic connotation is lost along with its non-standard basis (the implicature no longer having any reason to be triggered), or it overcomes this obstacle and eventually ends up being conventionalized.

In the later case, speakers would not have a clue why such a term comes with such a connotation. Here is a short digression about the former case, illustrating why and how autonymic connotation, and Nunberg's mechanism for STs, might in some cases resist conventionalization.

We just saw that among the inferences that we can draw based on metadata, there is the class of inferences arising from the fact that the term that is used is non-standard (corresponding to autonymic connotation). It is therefore sometimes because a term t belongs to a marginal group that a speaker can successfully present herself as marginal by her use of the term t.

This fact is not specific to language. For example, it is because the once marginal members of the hip-hop community were wearing their cap backwards that wearing a cap backwards could be used as a sign of marginality. This works until so many people want to present themselves as marginal with this sign that the sign itself becomes mainstream. When this stage is reached, wearing the cap frontward again can signal marginality. The sign of a backward cap is thus hard to conventionalize as a sign of marginality.

The same phenomenon is observed with touristic destinations. When one finds a beautiful and "authentic" destination, one tells others to visit the place. With more and more people visiting the place, its "authenticity" is soon replaced by a tourist directed economy. That is, what makes a place a good touristic destination vanishes as soon as it becomes a touristic destination. Conventionalization of x removes what makes x possible in the first place.

These analogies illustrate how autonymic connotation might resist conventionalization. But although we see why autonymic connotation could resist conventionalization (in accordance with Nunberg's point), we also see that this barrier is not absolute.

A backward cap might very well become the conventional sign of marginality, even though in the end everybody wears their cap backwards. Wearing a cap backwards would then naturally mean that one is a conformist and non-naturally mean that one is marginal. People wearing caps backwards would just be conformists presenting themselves as marginal. There is in that no contradiction.

Similarly, the touristic destination might very well stay for a long time a very famous and "conventional" touristic destination even though everything that made it a good touristic destination in the first place has vanished. Tourists would just be followers taking themselves to be explorers, and that is not an impossible state of affair either.

Going back to STs: autonymic connotation might constitute an obstacle to conventionalization. Maybe Nunberg is right about a subclass of STs whose negative associations have not yet been conventionally integrated in the content of the term. It is hard to know whether the connotation is conventional or not, but let us assume that it is possible that the connotation of the N-word has been conventionalized over time.

Recall that it is important for Nunberg's view that the negative aspect of STs is not part of any conventional dimension coming with it:

> On the utterance-condition view, it's conventional among English-speakers to use "nigger" to refer to blacks in order to express racist attitudes [...] On my view, roughly, it's a convention among certain English-speakers who have racist attitudes to use "nigger" to refer to blacks (Nunberg, p. 46)

But this does not exclude the possibility of a conventionalization mechanism making some of these negative aspects of STs part of their semantic meaning, in the format of an utterance-condition or of a presupposition for instance. I fail to see why all STs would resist standard mechanisms of conventionalization (such as the conventionalization processes of force described in Benveniste 1958 or Recanati 1981), and eventually have their enriched content integrated.

There surely are non-semantic sociological differences between terms having the same reference. It is a common fact that a term t is used by a social group g to refer to x, whereas a term t' is used by another social group g' to refer to x as well. But that such a social difference fully explains the difference between "African-American" and "nigger" seems implausible.

In the right context, we can certainly denigrate by using words that are not conventionally derogatory (e. g. "student" or "philosopher"), but it is no less obvious that certain words have acquired a conventional derogatory force. I am trying to account for these terms too.

It seems that the meaning vs. metadata distinction is to some degree arbitrary. We could maybe even capture all presuppositions in terms of metadata about "the situation where the term is used". It's not because we can construct a diachronic/pragmatic story that the synchronic/conventional bit is destroyed. Nunberg's view might be best seen as a reflection on the diachronic origin of potentially synchronic aspects of language, rather than a general, exhaustive and predictive account of STs in natural language.

When it comes to the synchronic, cognitive, psychological aspect of meaning, what we need to know is what exactly is needed to be competent with the term, and that is a different story.

To sum up, Nunberg's slogan could be: "If you want to know whether "Redskin" is offensive or not, look at who uses it and how they use it"[44]. It is a fact of every single term t that its meaning, in the broad sense, will always be a function of who uses it for which purpose[45]. It could be that Nunberg's deflationary account of STs is a consequence of a deflationary view of meaning in general, and does not in the end help us identify structural specificities of expressive terms like STs.

Nunberg's account, even under my three-steps reconstructed version, does not give necessary nor sufficient conditions for a word to be a slurring term. One may use a word to signal affiliation, and the word is not a slur; and the use of a slur may be dominant and offensive without there being a relevant alternative to that word.

[44] Note for example that numerous racist terms come from the police terminology: 'YBM' for young black male, "925" for suspicious person, "deuce" for black people because on the Philadelphia police form "2" is for black people, "DWB" for "driving while black", "nog" for "nigger out of gas", "slide" for blacks, "spliv" from washington DC area police, "trog" for unemployed whites etc. We can know that these terms are STs only on the basis of their provenance, the way they are used, for what purposes etc.

[45] B. Spector remarked that it is usually not so much about *who*, and that is why STs could be different (p.c.).

Chapter 4. Toward an Account of Slurring Concepts

The present chapter is a transition from purely linguistic (hybrid) accounts of slurring terms towards psychological accounts of slurring concepts.

First, I will develop the objection raised against most linguistic accounts of STs, stressing the need to account for another related, and possibly more fundamental, phenomenon: slurring thought. This will eventually lead us in the following chapters to attempt to analyze slurring concepts (SCs) rather than slurring terms[46].

Second, I make a few methodological remarks on my understanding of concepts, the goal of an account of slurring concepts in its relation with an account of slurring terms. This new focus on the conceptual level will lead me in the third section to update the initial explananda we started with, taking seriously the mental correlate of the phenomenon of slurring as the central aspect of the phenomenon.

I then introduce a general understanding of the distinction between central and parasitic uses, calling on Millikan's notion of *proper function* and *normal conditions* to make sense of this explanandum. This explanadum indeed has a special status because its explanantia is somewhat theory-dependent, and it shall thus be treated separately.

[46] Many thanks to Michael Murez, whose input was crucial in the development of this chapter.

4.1. From Communication to Thought

> There is much to be said for the old-fashioned view that speech expresses thought, and very little to be said against it. (Fodor et al 1974)

The investigations of STs in the previous three chapters reached an end, and it seems that they are incomplete. Some of them might be accurate, so far as linguistic matters are concerned, but there are other interesting issues apart from the linguistic issues that a philosopher interested in slurring representations might care about.

We saw that STs are pejorative terms that target groups on the basis of their ethnicity, gender, sexual orientation and the like, e. g., "kike", "nigger", or "faggot". We saw that such terms are usually taken to have hybrid semantic content, in the sense that in addition to purporting to pick out a worldly referent (a social category or group), they also somehow express or display negative emotions or attitudes of the speaker (e. g. Potts, Jeshion).

This expressive aspect of STs has the curious linguistic property of "projecting" out of most semantic operators like negation, conditionalization, modals and so on. For example, a typical utterance of "Chomsky is not a Chink" expresses negative attitudes towards Chinese people, despite the fact that the slurring predicate is under the scope of negation. So the expressive content of "Chink" is not affected by negation. Similar phenomena occur for other STs and other semantic operators.

In recent years, linguists and philosophers of language have taken a strong interest in the expressivity of STs, and in their seemingly closely related distinctive projective profile. In an effort to account for these phenomena, they have invoked a variety of notions from semantics and pragmatics, such as presuppositions (Sauerland 2007, Macia 2002, 2006, Schlenker 2007), conventional implicatures (Potts 2007, McCready 2010, Gutzmann 2015), conversational implicatures (Nunberg 2017), and speech-acts (Langton et al. 2012, Bianchi 2014b).

Although theoretical progress has been made in recent years, no consensus has been reached and I have shown that most current hybrid accounts are not fully satisfactory. However, my goal in the remaining of the present work is not to settle the debate regarding the alleged hybridity of STs, but to question one of its underlying assumptions.

STs have thus far been considered to be largely, if not exclusively, a linguistic phenomenon. Yet as the notion of "hybrid content" might already be taken to suggest, the phenomenon of STs is not merely a matter of speech, nor even of language.

Slurring terms characteristically express attitudes or mental states. A complete account of the phenomenon of slurring should therefore, I argue, include an investigation of its psychological component. If we want to fully understand slurring, we would do well to take a detailed look not only at slurring terms, but also at the psychological states of the subjects who employ them.

Here, I thus propose a novel conceptual approach to STs. According to this approach, which I will justify at more length in what follows, it is not simply because a certain rule of linguistic usage has been transgressed that STs have their characteristic properties i. e., that they are expressive, derogatory, offensive etc.

It is rather that their usage tends to reveal something about the speaker's distinctive conception of the social world. Uttering a ST is not merely a kind of linguistic *faux pas*, on a par with the utterance of other taboo words (such as e. g. "shit").

Rather, the choice to use a ST (at least in certain contexts) betrays that the speaker's private representations of social reality are defective, in a characteristic manner that is likely to cause offense.

The driving hypothesis of my conceptual approach is thus that a philosophical account of slurs should extend beyond STs, and take into account what I dub "slurring concepts" (SCs) - those concepts that are *normally* expressed by STs, in a sense which will become clear in what follows.

For example, Gérard, a Germanophobic Alsatian, might use the term "boche" when talking to or about his German neighbor. In so doing, Gérard betrays the fact that when he thinks of his neighbor, he categorizes them using the concept BOCHE.

Let me now argue a bit more precisely for the following hypothesis:

> **Slurring Concepts Hypothesis (SCH)**: Slurring is not merely a matter of speech. There are slurring concepts in thought (and Slurring Terms express such concepts).

SCH entails that a suitable account of STs shall not rely merely on speech, for the communicative aspect would be just one side of the coin. Independent of communication, racists have racist *thoughts* about their targets, which differ from neutral thoughts, and it is this kind of thoughts that I want to account for.

I thus claim that it is worth trying to derive the linguistic properties of STs, such as their offensiveness or their wide projection profile, from structural properties of the concepts they are used to express (that is SCs). SCH relies on the (common) assumption that thought is prior to language, and that expressions can inherit (some of their) properties from mental representations. The resulting view will not rely on essentially communicational mechanism to account for the phenomenon, and will thus not be compatible with merely linguistic accounts.

It seems that linguists and philosophers are interested in slurring terms for at least two reasons: first STs are bits of language with political and social import, so that an understanding of their functioning might help issues of the civil society to be better managed.

Second, as we saw earlier, the effects of STs are hardly reducible to the truth-conditional layer of meaning, and could thus contribute to a better understanding of the varieties and specificities of non truth-conditional meanings. I suggest that none of these two questions are satisfactorily answered by merely reducing of STs to use-conditions, as the main hybrid accounts we considered tend to do.

Beside the specific objections I have raised to some of the hybrid-theoretical accounts of STs, a more principled objection can be opposed to all of them. On each of the main three hybrid accounts I have discussed - the presuppositional account, the conventional implicature account and the speech-acts account - STs are a purely linguistic phenomenon, in that the expressive component derives from conventional constraints on the use of certain linguistic expressions.

But prejudiced representations are much more than a linguistic phenomenon. Can't people also *think* pejoratively when they think about their targets? Do we want to restrict the phenomenon to matters of language use? Maybe there is indeed something like a condition of use on these terms, but there is something unsatisfying in this conclusion, as it seems an essential question has not even been addressed.

These terms have the properties they have for a reason. What is wrong is uttering a slurring term is not merely that a convention of language was transgressed. What is wrong happens at a deeper level, it seems.

An utterance of a ST, in addition to showing that the speaker does not conform to certain conventions, reveals something about the speaker's inner world, about how she categorizes and articulates representations in a manner that is not ours, that we find flawed.

What is it exactly that is wrong in thinking of someone as a "Boche", as opposed to "German"? What is wrong is the Germanophic perspective on Germans, but what is it to have a particular "perspective" on Germans, how is that materialized, how does it work? These are the questions whose answers would give a more complete and satisfying answer to the problem of slurring terms.

The study of purely communicative matters is of course interesting. But I take the study of linguistic constructions as the indirect study of some dimension of our cognitive architecture. The goal is not to study the human communication system in itself, as it is not even clear yet that language's primary function is to communicate.

But since we happen to use language in communication, studying communication can be a good strategy to study language. If language is above all an internal tool to represent the world[47], properties of externalization merely gives us a hint about more fundamental properties of human cognition. In this sense, stopping an investigation of STs right after being convinced that they follow a rule of language use could be a bit like studying symptoms, without attempting to know more about the disease that produces them.

[47] This is a tool that we happen to be able to externalize (*via* gestures, recycling our gripping ability, or *via* sounds, recycling the breathing and eating system).

Mere use-conditional hybrid accounts of symptomatic uses of STs are not equipped to make room, at a deeper level, for slurring thought.

Let us start with the speech acts account. Speech-acts are, as their name indicates, a matter of speech. The relevant speech act, in the case of STs, is something like the expression of a derogatory attitude, or the affiliation with a group of racists for instance. STs are linguistic expressions which both have a descriptive meaning, in virtue of which they denote certain objects or people, and are conventionally used to express a derogatory attitude toward these objects or people.

Under any version of the speech act account, the two elements - the descriptive concept expressed by the word and the derogatory attitude also conventionally expressed - only combine at the linguistic level, in virtue of the conventions governing the expression: the expression has both a descriptive meaning and a certain "force"[48].

At the mental level, however, there are two separate elements: there are concepts that are purely descriptive and apply to certain objects, and there are positive and negative attitudes towards these objects (Frege, 1918).

Naturally, we should at least consider the possibility that some concepts, some mental representations, may be irreducibly colored with an attitude, at the mental level, independently of maters of linguistic expression. Maybe the combination of categorization and attitude is only made possible by language, but this has to be established and cannot be simply taken for granted.

[48] Bolinger (2015) for instance wants an account of STs to parallel a story for rude speech, as both categories can be "insulated" (i. e. plugged under direct quotation), project widely, are autonomous and susceptible to derogatory variation.

A consequence of my present argument is that, although both phenomena appear similar when one looks at linguistic and conversational data, this does not mean they stem from one and the same source: where rude speech triggers offense and projective meaning because of a deviation with respect to a linguistic social norm, STs trigger offense and projective patterns because of what they show about the speaker's mental architecture.

The same considerations apply to the account in terms of conventional implicatures. We have seen that the notion crucially involves the idea of a multidimensionality of meaning (with certain compositional rules under Pott's account), but how these dimensions should be characterized is still to be discovered. In order to show how the objection applies to this view, I will discuss what might be the best current account: the use-conditional account of CIs, that is, CIs as conditions imposed on the use of certain terms (Kaplan 2001, Recanati 2002b, Predelli, 2013).

Why does a CI term like "but" or "even" somewhat imply (but doesn't say) that, lets say, there is a contrast between the two conjuncts in the former case or that there is something unexpected in the later? An adequate way to conceive of it is to distinguish between truth-conditions and use-conditions. Under this story, a CI term *x* makes a truth-conditional contribution, and has also attached to it a rule of use of the form:

(116) Utter x only if conditions p and ... and q are satisfied in the context.

An utterance of a CI term, in virtue of the rule imposed on its use, implies that the conditions are fulfilled (or at least that the speaker thinks they are fulfilled, I don't enter in the details here). This is a "pragmatic" implication, that is, something that the use of an expression implies, without necessarily being directly encoded in the standard descriptive format. It might still be indirectly conventional, only in the sense that the rule is itself conventional.

Similar to STs, the contrast between an interjection like "oops" and its approximate truth-conditional counterpart "I've just witnessed a minor mishap" is, according to Kaplan's paper, due to the way the information (identical in the two cases) is couched (Kaplan, 2001).

Predelli equates the descriptive way with Kaplan's character and assimilates the expressive way with what he calls bias: a restriction on the class of contexts in which character is evaluated (Predelli, 2013), which can be understood as the formalization of a rule imposed on conditions of use.

Under these general approaches, semantic information can be conveyed under two formats: either by *truth-conditional*, descriptively encoded material, or by *use-conditional* means. I make the claim here that these restrictions on the contexts of use (i. e. rules imposed on the use) can model all other sorts of non-truthconditional meanings: presuppositions, vocatives, nicknames, interjections, honorifics, gender marking, register etc.

Mixing truth-conditional semantics (what is said, the relations between words and things) and use-conditional semantics (what is implicated, the relations between utterances and contexts) is indeed an interesting research program. If that is what's happening with the CI view of STs, then the phenomenon of derogation is clearly a linguistic one, based on linguistic constraints on the use of certain expressions.

This does not make room for the possibility of pejorativeness in thought. Why not? Because, if it is the use-conditional dimension that is responsible for the evaluative content conveyed by STs, then the descriptive and evaluative dimension in STs combine only in, or in virtue of, communication. That is not to say that a use-conditional theory makes wrong predictions, but I remark it is stuck at the social, communicational, level of language.

But what if the linguistic, communicational phenomenon we started with happened to be consequence of a deeper, non-communicational phenomenon? A theory might describe adequately a phenomenon without being at the most adequate level of description, and that is what I think is happening to most current hybrid accounts: it describes STs adequately and make correct predictions, but its scope is more limited than necessary.

It is for instance unable to explain why STs have the use-conditions they have, or how STs function when they are deployed privately by a speaker in the first person, independent of communication. As we shall see, there is room for a more general theory that would not be less parsimonious, explain the linguistic data, and keep room for other related (psychological) phenomena.

Rather than objecting to use-conditional approaches by claiming that STs are associated with such or such conventional condition on their use, I propose to develop a broader account of SCs that could eventually also explain why STs appear to be associated with such or such conventional condition on their use.

Appearances notwithstanding, the presuppositional account also falls under that objection. To be sure, presupposition initially corresponds to a mental attitude: that of "assuming" something or "taking it for granted" (Stalnaker, 1973). That is not a specifically linguistic phenomenon. But there is, in addition, a linguistic phenomenon: some expressions can be felicitously used only if the speech participants presuppose something in the more basic, mental sense.

Presupposition as a lexicalized feature of some expressions can be modeled in terms of use-conditional semantics:

(117) Utter x only if p is "assumed" or taken for granted in the context of use.

The presuppositional account of STs follows this pattern: the suggestion is that the conventions of language dictate that certain expressions (slurring terms) can be felicitously used only if the speech participants all presuppose (in the basic, mental sense) that the targets are despicable, or something like that.

Again, this locates the combination of the two ingredients at the linguistic level, and we should consider at least the possibility that the combination might already occur at the more fundamental, mental level. We should consider the possibility that some concepts may integrate, from the start, some emotional or attitudinal component. Indeed, I do not see what principle would prevent emotions from being part of concepts (see Richard, 2008).

There are several pathways linking emotional processing to high level processing (Bechara et al. 2000, Pessoa, 2014), and it is easy to see the evolutionary advantage of having a conceptual processing involving emotions (for action etc.). That is not to say that there is no paraphrase at all, but this fact can also be seen as a hint that the ineffable, emotional, non-conceptual dimension should be taken more seriously. That there is such a thing as prejudiced thought is a very likely possibility, and we don't want a theory of slurring to exclude this very likely possibility.

My main goal from now on is to provide an account of such slurring concepts. I hypothesize that they form a distinctive class of mental representations, whose psychological characteristics are interesting in their own right, and which also promise to help throw light on broader issues, such as the interplay between representational and non-representational dimensions of content.

Before evaluating different conceptual accounts of slurring, I shall make a few methodological remarks. First, it is worth distinguishing modest and radical versions of the conceptual approach to slurring.

> **Modest conceptualism** maintains that an account of slurs that does not take into account the underlying psychological states of speakers is *incomplete* i. e., that an account of slurs that takes into account underlying psychological phenomena has epistemic and theoretical virtues that an account that largely or exclusively focuses on language lacks.
>
> **Radical conceptualism** maintains, more stringently, that an account of slurs that does not take into account underlying psychological states is *incorrect*.

I mean to commit only to the former, modest conceptualist view. Linguists, for example, may be perfectly justified - e. g. by considerations having to do with the division of scientific labor - in remaining silent on the sorts of issues I focus on now.

I am thus happy to grant that an account of STs that is largely neutral, or even silent, on the psychological underpinnings of STs can be descriptively correct, so far as it goes. I simply maintain that such an account does not go as far as it potentially could i. e., that there is much to be gained by relating STs to their conceptual counterparts, and by viewing slurring as a phenomenon that spans the boundary between language and thought.

Note that I use the term "concept" as it is standardly used in psychology, to denote mental representations. Much of what I will have to say, however, should be easily translatable into terms that a philosopher who instead identifies concepts with *abstracta* would be happy to accept. Indeed, I intend to remain as non-committal as possible with respect to disputed issues concerning the metaphysics of concepts.

I assume that concepts are roughly "word-sized" constituents of thoughts or mental states, which are normally endowed with semantic content, and which normally serve to make reference to a category. I leave open the possibility that some concepts are empty, i. e., fail to successfully pick out any actual or even possible category of individuals.

I am neutral with regard to whether or not concepts have a certain amount of internal structure, so that it makes sense to talk of their various parts or constituents. I thus do not make any specific assumptions about the precise nature of the relevant constituency relation.

Again, much of what I will have to say about the internal structure of slurring concepts could be recast without loss in terms that a radical conceptual atomist would accept, e. g., by rephrasing my claims about the various constituents of a concept in terms of structured "conceptions" associated with atomistic "concepts".

I will return at some length to the issue of exactly how slurring terms and slurring concepts relate. However, I wish to immediately make clear that I do not assume a one-to-one correspondence between STs and SCs. I take as my starting hypothesis that what I call "slurring concepts" stand to slurring terms in a manner roughly analogous to the way that (tokenings of) first-person concepts stand to (tokenings of) first-person terms.

It is widely accepted that one can think of oneself as oneself independently of using, or even being able to use, the term "I". Just as it seems plausible that subjects may possess or deploy first-person concepts independently of actually using, or even being competent with, first-person terms, I take it that subjects can acquire, possess and deploy slurring concepts (to some significant degree) independently of uttering, or even being able to competently utter, slurring terms.

While I thus make room for the possibility of usages of slurring terms which are not accompanied by deployments of slurring concepts, my view is that normal uses of slurring terms do serve to express slurring concepts. As a result, I appeal to uses of slurring terms in language to help fix the reference of the theoretical notion "slurring concept". I introduce the technical expression "slurring concepts" to denote the sorts of concepts that such terms normally express.

Although the main focus of the remaining of this work is the nature of slurring concepts, i. e. a psychological phenomenon, I further assume, as a working hypothesis, that slurring terms in language inherit many of their most interesting properties from the concepts they are normally used, and normally taken, to express.

Thus, I hypothesize that the investigation of slurring concepts in thought will ultimately help to explain distinctive characteristics of slurring terms in language and communication. A secondary goal of an account of slurring concepts is thus to help advance our understanding of slurring terms.

I might therefore be fairly taken to have as a broader aim to account for the phenomenon of slurring representation in general, a phenomenon I take to encompass two logically and empirically distinct types of representations – slurring concepts and slurring terms. Because these two types of representation, though interestingly related, remain distinct, one could in principle accept an account of the former, while rejecting what I have to say about the latter (or *vice versa*).

My main goal now is to provide an account of slurring concepts based on the *prima facie* plausible hypothesis that they form a relatively unified, theoretically interesting class i. e. a "psychological natural kind". I will soon return to how I think of kinds of concepts.

Suffice it to note for now that I do not suppose that the kind "slurring concepts" possesses a sharp, classical essence - any more so than other candidate kinds which are targeted by special sciences like psychology. My aim is therefore not to provide necessary and sufficient conditions that a mental representation must meet to be included in the class of "slurring concepts".

Rather, having fixed the reference of "slurring concept" by appealing to normal uses of slurring terms, my goal is to throw light on the nature of such concepts by setting out a number of theoretically interesting, distinctive traits that such concepts are very likely to possess, in normal contexts - not by accident, but also not without exception.

4.3. Updated Explananda

Since I am now dealing not only with slurring terms but also with slurring concepts, which I take to be prior, I shall update the explananda that an account of slurring representations has to deal with. I thus repeat below the initial explananda we introduced in chapter 1, but this time adapted to the novel notion of slurring concepts. I apologize to the reader for the amount of repetitions in the six following pages, since a lot below still concerns STs, the expression sof SCs, but I have favored a complete updated list of explananda.

4.3.1. Updated Central Explananda

- *Hotness*. In many important cases, SCs seem to involve an emotional or affective states, or to feature an evaluative component that appears to be irreducible to their descriptive semantic value. Any theory of SCs should give an account of the nature of this intuitive "hotness" of SCs. A related question is how the hotness of SCs varies across thinkers (racists and non-racists). The expressivity of STs could be the linguistic correlate of SCs hotness.

- *Possession conditions*. A right account of SCs should account for which subjects count as normal possessors in which conditions, in a sense of "normal" that should be independently clarified. This is important, because it is clear that there are different classes of possessors of these concepts. There are the primary ones who possess and deploy SCs because they are racists, and there is the rest of us, who somehow understand STs but do not use them.

- *Defectiveness*. There is a strong intuitive sense in which SCs are flawed concepts. A theory of SCs should strive to locate their main defect(s). Here, I pursue the working hypothesis that SCs are not merely morally or ethically flawed, i. e. they also involve cognitive flaws, and correspond to an inappropriate way of categorizing and reasoning about reality.

Note that this explanandum has a very special status, because talk of "inappropriate" or "defectuous" ways of categorizing is normative, and in this sense does not have the scientific neutrality we expect from an investigation into a class of terms. I think it is worth giving pride of a place to it though, because as theorists, we all intuitively recognize that these terms are flawed, and clarifying in which sense exactly we take them to be flawed might be enlightening about their nature and specificities as linguistic entities.

- *Neutral Counterparts (and extension)*. We saw that STs were taken to have NCs. SCs shall then have their NCs too. This comes with two related explananda. The first is that a proper account of SCs should explain how they are alike, and how they differ from one another. Additionally, seemingly coextensive SCss such as KIKE and YID must differ in their cognitive roles. A proper account of SCs should individuate them so as to allow for that fact. The second explanandum that comes with the notion of NCs is to say what the extension of SCs is - leaving open the possibility that they have a null extension and simply share their "target" with NCs.

- *Dehumanization and identifying thinking*. Jeshion (2013) argues that one of the main role of STs is to encode dehumanizing modes of thought. SCs are good candidates to be such dehumanizing modes of thought. The cognitive act of dehumanization and its relation to slurring concepts should certainly be clarified. Dehumanization is clearly involved in racism, as can be seen in the common theme likening people to animals (Jahoda, 2015). See Haslam (2006) for an integrative review on dehumanization.

Deployments of SCs also seem to classify the targets so as to reduce their identity, as if being e. g. a BOCHE was what the target really *is* deep down (again, see Jeshion, 2013). Note that the identifying component of SCs seems linked to hotness, because it seems that thinkers succeed in identifying their targets through a hot negative attitude, such as regarding the targets as fundamentally inferior.

- *Derogatory variation*. Since some STs are more offensive than others (Hom, 2008), some SCs could be more "hot" than others. There is inter-group variation (for instance, the anti-Semitic KIKE seems to be more pejorative than the outdated Germanophobic BOCHE) and intra-group variation (for instance, the racist concept that the n-word expresses is way more offensive and pejorative than the somewhat rare SPADE, although the two concepts have the same target). An account of the hotness of SCs and of the offensiveness and expressivity of STs should thus allow for such hotness, offensiveness, and expressivity to come in degrees.

Cognitive role

Qua concepts, SCs play functional roles in cognition that should be clarified. What inferences do they license? What sort of inputs (experiences, perceptions etc.) trigger their deployment? What kind of actions are they prone to elicit as outputs? A few candidate cognitive roles of SCs:

- *Utterances* of STs are an output of SCs.

- *Categorization judgments* of groups and individuals as SCs trivially involve SCs.

- *Fineness of grain*. A proper account of SCs should explain how they differ from their so-called "Neutral Counterparts". Furthermore, coextensive SCs such as KIKE and YID differ in their respective cognitive roles. A proper account of SCs should individuate them so as to allow for that fact.

- *Understanding*. It is likely that SCs play a role in allowing most thinkers, even those that do not use STs nor possess SCs, to understand or access (at least part of) the meaning of STs. Hom & May (2013) take this question to be the central question about STs:

> How can a competent, rational speaker of a language know the meaning of a pejorative without being committed to, or even complicit with, racist attitudes? (Hom & May 2013, p. 1)

It is interesting that every competent of the language has the capacity to understand the terms even without being disposed, in any situation, to use them. Indeed, everything looks as if SCs were not part of the mental apparatus of non-racist thinkers, because they are not concepts they deploy, but they understand to some extent what the terms are about and perceive their offensiveness.

We shall thus distinguish between the meaning of SCs as possessed by "normal" possessors, and the meaning of SCs as possessed by the rest of us, because the two are not necessarily the same.

- *Ideologies and Stereotypes*. SCs seem to embody, or to be associated with, racist ideologies and stereotypes about the targeted groups. There are at least four independent reasons to believe that stereotypes play a non-negligible role in shaping SCs:

i) STs simply tend to bring stereotypes to mind (Jeshion 2013b).

ii) Stereotypes also dehumanize and harm the target's self-conception (Jeshion 2011).

iii) Stereotypes and STs are both associated with a taboo (Anderson & Lepore 2013)[49].

iv) Derogatory variation might indicate that different stereotypes are associated with different targets.

An account of SCs should be able to characterize the nature of the connection between SCs and ideological and stereotypical thinking about the targets.

- *Contempt*. SCs seem to be closely related to negative moral emotions, such as contempt. The nature of the link between deployments of SCs and the relevant negative moral emotions should be clarified. Do emotions intervene in categorization, and if yes, how exactly?

Connections with STs

If STs normally express SCs, we should expect a good account of SCs to contribute to the explanation of many interesting features of STs, such as:

- *Projection*. We saw that the fact that the expressivity of STs scopes out of most semantic operators is considered by many theorists to be the main property of STs to account for. As STs are now taken to express SCs, we shall aim at explaining this important linguistic property of slurring terms from properties of slurring concepts.

[49] Note that I take it to be unlikely that the prohibition imposed on STs is sufficient to account for all their features - as Anderson & Lepore (2013) seem to be aiming at. The main reason is that there exist languages with certain phonological forms that can express either a STs or another non-slurring term. For instance, the Italian "finocchio" is either a ST targeting homosexuals, or the name of fennel. What is it that could be prohibited in the former case and not in the later?

- *Offense/Derogation*. Uses of STs offend and derogate targets and bystanders. This is a simple observable feature of STs that shall be accounted for. Not all terms offend and derogate, we thus need to understand how certain terms seem to have acquired this particular power, especially if it derives from properties of the concepts they express (SCs).

- *Reluctance to evaluate*. Another property of STs is our reluctance to attribute truth (or falsity) to 3rd person descriptive statements where they are used. Our mitigated intuitions as competent speakers with regard to the truth and falsity of such racist statements is surely a symptom of a complex interaction between different components of our linguistic and social competences, and any account of SCs and STs has to offer an explanation for this striking piece of data.

- *Derogatory autonomy*. Another significant feature of STs is the apparent autonomy of their derogatory force from the beliefs, attitudes, or intentions of their users (Hom, 2008). Such an autonomy might not be the case of SCs though.

- *Various perlocutionary powers*. Slurring is a speech-act. In addition to classification and expression of negative attitudes, STs are also used to *do* different sort of things, to cause a variety of harm to the targeted groups.

We can try to identify these action potentials that STs seem to carry with insulting, denigrating, humiliating, stereotyping, belittling, dehumanizing, assaulting, making propaganda, subordinating, affiliating oneself with a group and so on, but such a precise description of complex acts one can perform with the *things* these words are is beyond the scope of the present work.

Nevertheless, accounting for STs should at least give us a hint about the way in which they come to have these different perlocutionary powers, and which of these come from properties of SCs. After all, these powers are also reasons why some speakers use STs.

Central and Parasitic Cases

Uses of STs/deployments of SCs come in large variety, and not all of them are on a par. Some are central and others are incidental. There is indeed a variety of special deployments of SCs or uses of STs that we need to account for. I follow Jeshion's pervasive identification of such uses and the terminology she introduced (Jeshion, 2013a).

First, some uses of STs seem not to display any offensiveness (call these *non-weapon uses*), such as *appropriated* uses among targets. Appropriated uses are uses of STs by the members of the targeted group as a friendly way to call each-others - "queer" and "nigger" have today such uses.

Second, some deployments of SCs target a sub-class of the category denoted by their neutral counterparts (g-contracting) and other uses are targeting a larger class than the neutral counterpart (g-extending):

Standard deployment of SCs refer to the same category of people as the neutral counterpart - call these *G-referencing* uses - and my discussion will be mostly focused on uses that are considered to be central: *G-referencing weapon uses/deployments* of slurring and concepts, in order to better examine the problematic relations between their truth-conditional - i. e. descriptive - properties and their potential to perform acts of offense, insult, moral evaluation etc.

The task of pulling these two sorts of uses/deployments apart should be distributed among i) the necessary preliminary work of isolating the phenomena to be explained and ii) the explanation of the phenomena.

This explanandum indeed has a special status because its explanantia is somewhat theory-dependent, and it shall thus be treated separately. I will be doing so in the next section.

4.3.2. Updated Peripheral Explananda

- *Creation and evolution*. It appears to be very easy to create STs and SCs contextually. Simply name the target group with the name of their favored food for instance, and you have a SC. Why and how do terms and concepts (so rapidly) come to be STs and SCs?

And why is their expressive power so sensitive to change over time, as shown by e. g. the amelioration of archaic STs like "boche" or "kraut"? Many STs were far more offensive in the past than they are today, and inversely, many STs which are extremely offensive today were arguably not that harsh just a handful of years ago. The meaning of non-offensive terms like "table" does not seem to change so rapidly, so why would STs be so sensitive to time?

- *Group formation, binding, and identity*. Uses of STs seem to play a non-negligible role in shaping communities. A trivial observation shows that sharing contempt for out-group members usually contribute to form and bind groups together, and STs and SCs could very well be used to play such a role.

- *Formation and perpetuation of social hierarchies, and of bigotry*. Jeshion (p. c.) argues that the expression of SCs as STs is also a device for building social hierarchies and for maintaining preexisting ones. They would also be useful to communities for the transmission of contempt toward the targets, hence for the perpetuation of bigotry.

- *Community endorsement*. Uses of at least some STs signal the community's endorsement of the bigotry expressed, it overtakes the speaker herself. It is not only the speaker's contempt that is expressed by uses of STs. In some cases, as we clearly see with the N-word, the harm and threat that represents a use exceeds whatever a single individual could do. It is as if uses of the N-word signaled that the community endorses the bigotry and oppression that the speakers expresses. Note that on top of bringing about the endorsement of the community, it seems that uses of STs also seem to bring to mind histories of past oppression.

In what follows, I will present and critically examine different candidate accounts of slurring concepts. I will assess each account primarily in terms of whether or not it provides adequate explanantia for all of the above. Before doing so, I start by addressing the last of the central explananda, which has a special status.

4.4. Understanding the Central/Parasitic Distinction

It is pre-theoretically clear that not all uses of STs and deployments of SCs have the same status. Intuitively, an in-group third-personal direct use of the n-word is a more central and paradigmatic kind of case than the appropriated "friendly" uses of "nigga" among targets.

Although a complete theory of slurring representations should account for all cases, and to say that a case is "parasitic" is not to say that it is unimportant or uninteresting, not all cases should be treated on a par theoretically.

We thus need to start by drawing a distinction between central and parasitic cases in order to know which data all accounts of slurring concepts need to handle. I propose to reanalyze the central/parasitic distinction in light of Milikan's notion of *proper function* (Millikan 1984, 1989):

> A proper function of [...] an organ or behavior is, roughly, a function that its ancestors have performed that has helped account for proliferation of the genes responsible for it, hence helped account for its own existence. (Millikan 1989, p. 289)

A similar notion of proper function can be extended to representations, as Millikan herself has argued. In the case of SCs, the proper function will be the function that is responsible for its creation and proliferation. I hypothesize that the central cases are the cases in which SCs accomplish their proper function. I call this hypothesis the **Central Cases Definition (CCD)**.

A consequence of CCD is that what makes something a SC is not a set of intrinsic features but a causal historical relation to its ancestors. My aim is therefore not to provide necessary and sufficient conditions that a mental representation must meet to be included in the class of SCs.

One could artificially provide necessary and sufficient conditions of the form: "being appropriately causal-historically related to such-and-such cases", but these would not be particularly illuminating. The important theoretical work lies in describing the proper function of slurring representations.

Thus, having fixed the reference of "SC" by appealing to normal uses of STs, my goal is to throw light on the nature of such concepts by highlighting a number of theoretically interesting, distinctive traits that such concepts are very likely to possess, in Normal contexts (in a sense of "Normal" defined below) – not by accident, but also not without exception.

So we have a theoretical understanding of the distinction between the cases in which SCs accomplish their proper function, and those in which they do not. But in practice, how can we pull the two apart? I propose the following heuristic:

> **Asymmetric Dependence Heuristic (ADH)**: Kinds of ST-uses/SC-deployments that plausibly stand in an asymmetric existence-dependence relation to other ST-uses/SC-deployments are (likely to be) parasitic.

For instance, there could be no "friendly" appropriated uses of the n-word without there being an offensive insulting use of it, whereas there are plenty of derogatory uses of STs without the targeted communities having appropriated them. The ADH seems to work for reclaimed uses: they must be parasitic.

Based on considerations of statistical frequency, one might further conjecture that the central ST-uses/SC-deployments are the expressive/hot third-personal ones, whereas second-personal uses are parasitic.

However, this is a somewhat risky inference, as proper functioning is not, in general, statistically most frequent. An argument would be needed to suggest an asymmetric existence-dependence relation between these two types of uses.

The notion of a proper function is closely linked to that of normal conditions. If the proper function of a SC is the function responsible for its creation and proliferation, then there must be conditions under which it was created and is able to proliferate. These conditions have a special status, then, because they are the conditions ideally suited for the emergence and proliferation of SCs. More generally:

> **Normal Conditions**: The normal conditions of emergence of an item are, roughly, the external (E) and/or internal (I) conditions under which the fulfilment of their proper function allows for its reproductive success. In the case of SCs, I consider two sorts of conditions: E-conditions i. e. social conditions; I-conditions i. e. psychological conditions.

I consider now two possible E-conditions for the emergence of SCs. First there could not be SCs without there being at least two groups, with at least one conceiving of the other as an "outgroup".

Second, intuitively, if the outgroup was regarded with sympathy or neutrality, the representation for it would not count as an SC. There must therefore be some kind of animosity, or at least competition, between the two groups.

I now consider and reject one possible I-condition of emergence for SCs. A common claim in the literature (Nunberg 2017, Bolinger 2015) is that slurring representations depend on the existence of NCs. Some authors go so far as to make NCs a necessary condition for slurring representations. That is what I call the Neutral Counterpart Constraint, which I discussed and rejected earlier:

> **NC Constraint (NCC)**: Slurring representations only emerge when they have NCs. There are actually four distinct subversions of this constraint:
>
> i) STs require NC-Terms
>
> ii) STs require NC-Concepts
>
> iii) SCs require NC-Terms

iv) SCs require NC-Concepts

We saw earlier that i) and ii) were to be rejected, because there are numerous exceptions - that is, cases of STs without an NC - such as "dark-skinned", "gook" etc. They can also be rejected on the grounds that we can conceive of STs targeting members of an out-group without an alternative "neutral" concept/term being available.

iii) and iv) do not leave space for stages in concept acquisition where children come to possess the SC and still lack the NC-C. They entail that one cannot acquire a slurring concept unless one already possesses an NC concept, which - as we saw earlier considering the case of a child coming back from school - is not really plausible. I prefer to reject the NCC constraint altogether, but nothing in what follows hinges on this decision.

Now that the framework for talk of slurring concepts is in place, let me investigate more closely the notion of hybridity we started with, and see how it applies or does not apply to slurring concepts.

Chapter 5. Calling Hybridity Into Question: Introducing T-Terms

The present chapter is an attempt to develop a first view of slurring concepts, by questioning one of my starting hypotheses: the idea that slurring terms and concepts are hybrid, that is, that their semantics contains two dimensions - one descriptive and one evaluative. The point of the chapter is to investigate whether slurs conform to the S-term model or to another model I put forward: the T-term model.

I start with a few remarks on the very notion of hybridity and its origins in Frege's notion of "tone". I then distinguish two classes of terms: S-terms, which are authentically hybrid, and T-terms, which are not.

Contrary to S-terms, T-terms are not co-extensional with their counterparts and have a richer descriptive content. They refer to a subclass of the group their counterparts refer to. The introduction of such terms helps me put forward a novel account of slurring terms and concepts: a reference-based account of evaluation (RBE).

This view argues that T-terms are, appearances notwithstanding, not truly evaluative: they simply have a rich descriptive content such that they refer to subgroups, subgroups which are independently, extra-semantically evaluated as being negative.

The evaluation ends up being associated with the terms, but it becomes associated only extra-semantically. The semantics of these terms is one-dimensional. I connect the debate to an existing literature on so-called "thick" terms and concepts, which raise similar questions having to do with two potentially separate dimensions of meaning.

I discuss the tension between the idea that the two alleged dimensions in these concepts are inseparable, and the so-called "objectionable" thick concepts in which the two dimensions appear to be clearly separate.

I then investigate the possibility that slurs are governed by RBE, that is are T terms. I address a series of objections to the resulting view of slurs. The first two are weaker and responded to, the last two will be decisive and lead us to investigate yet another view of slurring concepts. Where our reference-based view of the evaluative component of slurs did not locate the evaluation itself in their descriptive content, but derived it extra-semantically, the view I will explore in the next chapter will locate all dimensions of SCs, including the evaluation itself, in their truth-conditional content.

> Certain names carry two ideas; one, that we shall name "main idea", represents the denoted object, and another, that we could name "ancillary", represents the object as tinted with certain properties. For instance, the word "liar" means someone who did not speak the truth, but on top of that, it shows that we judge that the person we blame is mean, is maliciously hiding the truth, and thus worthy of hatred and despise. (Lamy 1678, 3rd edition, Book I, chap. VII, p. 24. [my translation])[50]

5.1.1. Frege's View on Hybridity

The idea that certain terms carry a species of semantic content in addition to their standard descriptive content was articulated long ago, as the above quotation shows. In more recent times, it can be traced back to Frege's view on differences between co-referential terms. I will start by clarifying Frege's positions on the objective and subjective dimensions of meaning. On the one hand there is *sense [Sinn]*, which corresponds to what is relevant in determining the truth or falsity of a sentence (Frege 1893). Sense is supposed to be purely objective and independent of psychology. Frege proposes the following analogy:

[50] Original (slightly adapted to modern European French):

"Il y a des noms qui ont deux idées ; celle qu'on doit nommer l'idée principale représente la chose qui est signifiée, l'autre que nous pouvons nommer accessoire, représente la chose revêtue de certaines circonstances [*sic*]. Par exemple, le mot "Menteur" signifie bien une personne que l'on reprend de n'avoir pas dit la vérité, mais outre cela, il fait connaître que l'on regarde celui à qui l'on fait ce reproche comme une méchante personne, qui par une heureuse malice a caché la vérité, et qui par conséquent est digne de haine et de mépris." (Lamy 1678, 3ème édition, Livre I, chap VII, p. 24)

> Somebody observes the moon through a telescope. I compare the Moon itself to the reference; it is the object of the observation, mediated by the real image projected by the object glass in the interior of the telescope, and by the retinal image of the observer. The former I compare to the sense, the latter is like the idea or experience. The optical image in the telescope is indeed one-sided and dependent upon the standpoint of observation; but it is still objective, inasmuch as it can be used by several observers. At any rate it could be arranged for several to use it simultaneously. But each one would have his own retinal image." (Frege, 1892, p. 30)

Frege insists on such an objectivity of sense so as to avoid having to relativize truth to human beings. Because sense is what links a concept[51] to its referent, it is also what gives propositions their truth-conditions (and truth-values when evaluated).

Surely, Frege thought, some propositions must be true eternally and immutably. For instance, it must be the case that 2 + 2 = 4, or that the earth is smaller than the sun, even in past and future times where humans aren't around. That the earth is smaller than the sun is not merely true-for-us, nor true-now, it is true *simpliciter*. And even if the sense of the term "triangle" appeared to evolve over time - as we see with many expressions diachronically - it is in fact not that the sense itself was modified but rather that the term became associated with another immutable sense.

The ontological status of sense has been and is still the object of extensive debates (see e. g. Dummett, 1973). Many commentators consider the "third realm" of immutable sense (which is neither physical nor psychological) ontologically problematic, but here is not the place to discuss this issue.

Because the relation between an objective sense and reference is truly objective and independent of the psychology of (human) subjects, it follows that sense requires no *thinker*: the existence of human psychology is not a necessary condition for the existence of sense.

[51] Here I do not use Frege's notion of "concept", but rather the notion of "concept" as it is standardly used in today's psychological literature, closer to the notion of a particular mental representation than to the sort of *abstracta* Frege had in mind.

Nonetheless, human beings would have the ability to grasp these objective senses, to entertain or deploy them in thought, and eventually to communicate them to one another through language. In virtue of their objectivity and of our ability to grasp them, senses can become public and communicable. The objectivity of sense is thus (i) what secures the reference of expressions and the truth-conditions of sentences in which they are involved, and (ii) what makes them public and communicable.

On the other hand, there seems to be bits of meaning that do not affect truth or falsity. As Frege and Grice pointed out, the difference in meaning between "and" and "but" might well belong to that second realm of meaning (Frege, 1879): "Mary is rich and honest" expresses the same thought as "Mary is rich but honest", but it also somewhat suggests that - simplifying - Mary's honesty is unexpected given her richness.

For Grice, as we saw earlier, this contrast is based on the utterances' *conventional implicata* rather than on *what is said* (Grice, 1975): both utterances, strictly speaking, *say* the same thing, but only one of the two conventionally implicates the unexpectedness relation between the conjuncts.

In Frege's terminology, we face here a difference of *tone* [*Beleuchtung*] (or *coloring [Färbung]*) rather than of sense. The same is true of sentences obtained by substituting "dog" for "cur" or *vice versa* (Frege, 1879, 1891, 1892a), which have the same senses but different tones. Both expressions "cur" and "dog" refer to dogs in virtue of their senses, but one of them conveys - simplifying - an additional depreciation of its referent.

Contrary to sense, tone is taken to be inherently *subjective*: it does require a bearer, in the sense that the existence of human psychology *is* a precondition for the existence of tone. Of course, in some sense, sense does require a bearer too: the bearer of sense is the expression. But it seems to me that Frege is saying, in the case of tone, something stronger than merely that tone requires bearers. Tone seems to require not only bearers but human subjects, thinkers.

A "naturalized" version of Fregeanism which would deny the existence of a third realm could claim that senses cannot exist independently of the expressions that have those senses, or even that they cannot exist independently of human minds. But there would still be, it seems, a difference with tone, which is subjective and mind-dependent to a stronger degree.

Dummett remarks that Frege's notion of tone is not as clear and intelligible as that of sense:

> What is not immediately clear is how it may come about that an assertion may be incorrect in any other way than by being untrue; how we may convey by what we say more than we are actually stating to be the case. (Dummett 1973, chap. 1, pp. 2-3)

That is, how could there be meaning where there is no description of a state of affairs? What is it that a statement could convey, on top of depicting ways that the world could be like? In order to better understand Frege's view on the nature of these aspects of meaning, consider different mental images that a florist and a poet might associate with the term "rose".

The florist and the poet both grasp one and the same concept of a ROSE (applying to roses), but differ as to the images they mentally associate with it[52]. In this case, the images are totally subjective and thus require a thinker. Only a thinker who entertains or deploys the concept in thought can add to it such personal and somewhat arbitrary associations. It is because they need a thinker that these associations are relative, and that they vary across speakers and (possibly) contexts.

Frege, who was primarily concerned with knowledge, gave a crucial role to senses in his system. The relations between senses are then characteristically rational/logic, and hence objective and non-arbitrary. On the opposite, mere "ideas", or mental associations such as the one of the poet and the florist, were irrelevant to knowledge. Frege thus draws a clear-cut distinction between the relations between senses, which are characteristically inferential, and those between ideas, which are characteristically associative.

Frege characterizes the differences between "cur" and "dog" in the following way. First, "dog" and "cur" are co-extensional. The difference between the two thus has to do with the ways they respectively act on the imagination of subjects who entertain them: "cur" is associated with negative attitudes and emotions that "dog" is not associated with.

As a consequence, the additional bit of meaning that "cur" possesses is as logically irrelevant as the additional bit of meaning that "rose" might possess when entertained by a poet/florist:

[52] This point is debated though. Wiggins (2016) for instance suggests that "horse" and "*Equus cabalus*" (scientific term) have the same reference and express the same concept, but have different senses.

utterances obtained by substituting one to the other have the same truth-conditions and express the same thought.

When evaluating the proposition that "This cur has four legs", speakers associate some negative feelings and attitudes with the expression "cur", but whether or not the dog referred to satisfies or justifies these negative attitudes does not affect the truth-value of the utterance. In that sense, the evaluative content associated with "cur" belongs to the same subjective realm as the poet's imagination.

But there is an important difference between the free and somewhat arbitrary associations coming with "rose" and the more constrained and shared negative content associated with "cur". Indeed, if the associations linked to "rose" are relative to subjects and arbitrary, those of "cur" seem conventional and stable across speakers (and contexts).

Note that I here oppose arbitrariness and conventionality. Since conventions are sometimes said to be "arbitrary", I shall clarify the sense in which I mean that mental images are arbitrary and not conventional.

When a mental association is subjective, it is "arbitrary" in the sense that it is up to each subject, depending on her particular perspective etc. This is not the kind of arbitrariness that applies to conventions.

Conventions are arbitrary in the sense that all that is needed for linguistic cooperation is to agree on a convention or another, but the particular choice of the convention is not important. That we say "hello" or "hallo" to greet each-other is not important as long as we all agree on the convention: the convention is hence "arbitrary".

Anyway, the difference between the "arbitrary" mental associations of "rose" and the more "conventional" negative associations of "cur" suggests that, even though the evaluation associated with "cur" requires a thinker and belongs to the subjective realm of associations, it differs from the images associated with "rose" with respect to the important aspect of communicability.

5.1.2. Disentangling Subjectivity and Communicability

Two distinct dimensions, subjectivity and communicability, might be better kept separate, where they were conflated in Frege's landscape. There might be, as Dummett puts it, a

> false dichotomy between mental images as subjective and incommunicable, [and] sense as objective and communicable (Dummett, 1973, p. 158)

That is, that something can be at the same time subjective and communicable is not a theoretical possibility for Frege. How could mental images, that are subjective, be communicable for instance?

One can draw here a simple type/token distinction to allow for that possibility, just like the type/token distinction is a way of understanding the communicability of sense. Tokens require concreteness, and hence might require a thinker. But types are abstract and mind-independent, they do not require a thinker.

Similarly, there might be communicable mental image types, even if tokens are concrete particulars and are hence not shared. Simply contrasting, as Frege does, senses with mental images is not entirely satisfactory: one must still explain the stronger sense in which images are supposed to be subjective.

We should thus clearly distinguish between two dimensions: the subjective/objective dimension, and the private/public dimension. Based on this distinction, the mental associations coming with "rose" could be both subjective and private. At the level of types, mental images are not particularly subjective.

At the level of token, they are subjective because they do not contribute to the truth or falsity of sentences and thoughts, and they are private because they are up to the thinker themselves: there seems to be no particular constraint on the association, hence they greatly vary across subjects (and possibly across contexts).

On the contrary, although the associations that come with "cur" are also subjective, they seem to be public rather than private. They are subjective because they do not contribute to truth-conditions and require a thinker, but they are not exactly up to the thinker. There indeed seems to be some conventional constraints on the associations.

In particular, it seems that the associations that come with "cur" *must* be negative in some way. Paradoxically, terms like "cur" whose evaluative content seems stable across speakers show that a piece of meaning can be, at the same time, subjective and conventional. Being interested mainly in the objective aspects of meaning, Frege does not really dig further into that potential tension between the subjectivity of tone and its apparent conventionality.

Frege's idea of separating two kinds of semantic components is today widely accepted, from Grice to the modern hybrid-expressivists *à la* Potts (for whom there are no differences in sense between an evaluative concept and its *neutral counterpart*, although there are differences in another conventional, *expressive* dimension). It is that consensual position, "hybrid expressivism"[53], that I will try to call into question in this chapter.

For expository purposes, it will be useful to separate such hybrid accounts in two distinct (although related) theses, both directly following from the very idea of a neutral counterpart. Note that the following presentation of hybrid expressivists account is slightly different from our earlier HEA in that I replace the notion of "co-extensionality" with that of "co-description"[54].

This version of hybrid expressivism is therefore not necessarily committed to the co-extensionality thesis (CET). I also use the phrase "evaluative concepts" to coin the class of concepts I am trying to clarify, so as to stay as neutral as possible with regard to their nature.

Here are the two sub-theses of the version of hybrid expressivism that I will question:

> **Co-Description Thesis (CDT)**: Evaluative concepts have the same reference-fixing, descriptive content as that of their neutral counterparts.

[53] Recall:

> **Hybrid Expressivist Accounts (HEA)**: Hybrid expressivist accounts of STs subscribe to the Co-Extensionality Thesis (CET) and call on other dimensions of meaning to account for their additional expressive properties.

[54] Compare the Co-Description Thesis (CDT) with the Co-Extensionality Thesis (CET) in the glossary p. 365.

> **Conventionality Thesis (CT)**: The evaluative component of evaluative concepts is just conventionally (hence arbitrarily) associated to their reference-fixing, descriptive component.

These two theses are connected to the two main properties we ascribed earlier to S-terms: being *hybrid* and being *separable*. CDT and CT indeed presuppose that "evaluative concepts" are hybrid, in the sense that they have meanings of two different kinds, for it says it has an evaluative component and a descriptive component. CDT also supposes that the two kinds of meaning it ascribes to evaluative concepts are separable, because it involves the notion of "neutral counterparts", which are objects sharing one but not the other type of meaning. And for a concept to be able to have one but not another type of meaning, the two types must be separable.

From this way of putting it, it follows that there are two ways to challenge Frege and others' hybrid expressivism. We can object to CDT by ascribing to evaluative concepts a richer descriptive content than that of their counterpart - including a stereotype and/or an evaluative component - or we can object to CT by introducing a real causal link between the descriptive and the evaluative components, thus going beyond a mere conventional or arbitrary association. I will consider both strategies successively, introducing T-terms and concepts.

I will now introduce another class of relevant terms, T-terms, in addition to our earlier S-terms, so as to investigate whether slurs conform to the S-term model or to the T-term model. Let us first call CDT into question. That is what Nunberg does, remarking that it is in fact not so clear that S-terms and their so-called counterparts are co-referential to start with[55].

Take the term "cur" for instance. It could be considered to be a S-term, with "dog" as its neutral counterpart and conventionally conveying a negative evaluation on top of that. But on closer inspection, it is not that clear that "cur" and "dog" are co-extensional. Consider the following definition of "cur" found in an online dictionary:

> A cur is a dog that isn't very good - or is a mixed breed. If dogs understood English, they would be offended at being called a cur. (vocabulary.com)

According to that definition, "cur" could well refer to a particular subclass of dogs, maybe to dogs that are dirty, or mean, or hostile, or a combination of some properties along these lines.

Similarly, the French noun "guimbarde" (just like the English "jalopy"), although it could seem at first glance to be a S-term referring to cars and expressing some sort of a negative attitude towards them, might as well be taken not to refer to the whole class of cars but to only cars that have certain additional properties.

The English equivalent "jalopy" does not really apply to any sort of a car either. One can certainly refer to a brand new Ferrari using the term "car" in English or "voiture" in French, but surely not using "jalopy" or "guimbarde" (disregarding ironic or other non-serious uses).

It seems rather that "guimbarde" and "jalopy" apply only to old and broken cars for instance, as indicated by the following definition:

[55] p. c.

> A jalopy is an old car that isn't working very well. You'd never call a new smooth-running car a jalopy. (vocabulary.com)

So both "cur" and "jalopy", which might have first seemed to be S-terms, could in fact refer to a narrower class than their so-called neutral counterparts. This would entail that their descriptive content is somewhat "heavier" than hybrid expressivists *à la* Frege take them to be.

Based on this observation, I will now investigate the possibility of building on a rejection of hybridity by re-assessing the co-description thesis (CDT). In order to do do so, let me define T-terms:

> **T-terms and T-concepts**[56]: terms/concepts targeting certain *subgroups* in virtue of (*i)* a reference to (at least) the *supergroup*, and (*ii)* an additional *descriptive* element[57] motivating (*iii)* an evaluation.

This definition involves new terminology, so I'll briefly elaborate. The idea is that T-terms are basically just like our everyday terms expressing everyday concepts. They refer to groups or individuals in virtue of (at least part of) their descriptive content, period.

They become more interesting only after one notices that the individuals they refer to all belong to a broader kind, and that there is or might easily be another term in the language referring to this larger kind.

One shall therefore distinguish between a *subgroup* - the reference of T-terms - and the *supergroup* - the reference of their (potential) alternatives. T-terms refer to the subgroup by adding some additional descriptive content to that of their alternative. So T-terms, as opposed to S-terms, are not co-extensional with their counterparts.

[56] "T-terms" for short.

[57] For comparison, recall the definition of STs (which is here not restricted to the specific subcase of slurs):

> **S-terms**: terms whose meaning is hybrid (it is made of at least two different kinds of meaning) and whose meaning components are separable (one can find or construct neutral counterparts)

I will still talk of their "counterpart" though, just keep in mind that the relation between T-terms and their counterpart is not one of co-extension, contrary to STs. I will now discuss the exact nature of the additional *descriptive* element.

According to an analysis of "cur" as a S-term, the term refers to dogs and is conventionally associated with a certain negative evaluation towards them. But now, according to an analysis of "cur" as a T-term, it simply refers to a certain subclass of dogs, and the question of the evaluative content seemingly coming with the term is open. So the current suggestion is that "cur" and "jalopy" are in fact not S-terms but T-terms. This will help us investigate whether slurs are T-terms rather than S-terms.

T-terms, if they exist, could give us a way out of hybridity and separability at one and the same time. T-terms are no longer hybrid in the sense that their content is just plain vanilla descriptive content, they express usual concepts like "tiger" and "table", and do not possess any sort of mysterious category of content like the "expressive" on top of that. The source of the observed "expressivity" will have to be found elsewhere.

Being non-hybrid, the question of the separability in T-terms does not even arise, trivially. One can wonder whether two dimensions of content are separable only when there are two such dimensions, but T-terms are one-dimensional.

The notion of T-terms also challenges the co-description thesis (CDT) and the conventionality thesis (CT) at once. Remarking that "cur" might refer to a narrower class than "dogs" is in itself questioning the co-extensionality thesis. And how can we derive the observed expressivity of terms like "cur" and "jalopy" then, if they just express standard descriptive concepts? Why does "cur" seem to differ from "dog" in its negative evaluative content?

The idea I will pursue now consists in introducing a link between the term's extension and the evaluative judgment. That equates calling into question the conventionality thesis. Where the conventionality thesis merely arbitrarily connected the concept with a negative evaluation, T-terms derive their evaluative import precisely *because* of properties of their reference. The evaluative content will thus not be arbitrarily or conventionally associated to the concept, it will be motivated and explained by its descriptive dimension.

The notion of an "additional descriptive component" is a little too underdetermined as it is and is compatible with different possible precisifications, as we shall now see. So what could be the nature of the additional descriptive element in T-terms motivating the observed evaluation coming with "cur" and with slurs?

What I want to do now is to gauge a novel simple position according to which that additional descriptive component in T-terms and concept has the nature of a standard reference-fixing component. We have already investigated other theoretical options earlier (e. g. assimilating the evaluative content to presuppositions, conventional implicatures or conversational implicatures).

So my current question is: can the evaluative content of slurs - understood as T-terms - be motivated solely by a *richer* reference-fixing component? The first alternative view to Frege's and other hybrid expressivists'[58] I will consider is thus the following view:

> **Reference-Based Evaluation (RBE)**: (i) T-terms express concepts with a descriptive component *richer* than that of their counterparts; (ii) T-terms and concepts refer to "bad" *subgroups*; (iii) Their evaluative content is (extra-semantically) associated with the perceived negative properties of the subgroup.

This view challenges both the co-description and the conventionality theses (CDT and CT) of hybrid expressivists *à la* Frege. The view has three tenets. First of all, RBE claims that T-terms express standard descriptive concepts, that is, it does not take the evaluative component to be an essential property of that kind of terms and concepts.

Second, and crucially, RBE claims that T-terms refer to *bad* subgroups. This second tenet is what will make the third, about evaluation, possible. Instead of saying that "jalopy" refers to

[58] Recall:

> **Hybrid Expressivist Accounts (HEA)**: Hybrid expressivist accounts of STs subscribe to the Co-Extensionality Thesis (CET) and call on other dimensions of meaning to account for their additional expressive properties.

cars and is conventionally associated with a negative evaluation towards them, "jalopy" refers only to a subclass of cars, for instance (simplifying) to old and broken cars. That is their semantic (and only) content.

Note again that it seems accidental that the bad members of the group are a just a subclass of the whole group, rather than the entire group itself. Another view seems possible that would stick to the co-extensionality thesis but still not be hybrid in the sense that the term would refer to the whole group while representing it descriptively as bad. I will focus on such a possibility in chapter 6.

Now because old age and brokenness are not exactly properties that we value in cars as a default, the reference of "jalopy" is itself *bad*.[59] That the objects referred to are themselves bad/good is crucial for the third tenet of the view.

In a nutshell, RBE is the view that T-terms and concepts are just standard descriptive terms and concepts which, given their reference to a *bad* subclass of their counterparts, and given that members of this subclass are *independently* evaluated, are naturally (non arbitrarily) linked to negative evaluation of their referent.

Note, importantly, that RBE is not drawing any semantic or inferential connections between descriptive content and evaluative content. This would raise many metaethical issues about the fact/value distinction.

Rather, RBE stays neutral on whether values can or not reduce to facts, and simply requires that the evaluation is performed independently of the content of the term. The term is simply descriptive and has a reference, period. The central element of the view is that evaluation is extra-semantic. The evaluation is then compatible with various views of how exactly the evaluation follows from the descriptive (semantic) component. I do not need to take a stand on such issues, as long as it is clear that the evaluation is extra-semantic.

[59] It is harmless here to assume a form of moral realism for the sake of clarity. If bland moral realism is not to the reader's taste, he or she could instead assume that the property referred to tends to provoke a negative response in human beings under normal circumstances, or some dispositional story along those lines. I here call such properties *bad/good*, as a shortcut.

Suppose for example that old and broken cars *are* bad. I do not intend to say much about what being bad for a car means, as all RBE requires here for the account is that the evaluation comes independent of language. So simply assume that old and broken cars are bad. Maybe that we naturally tend for some reason or another to dislike and disvalue cars that happen to be old and broken?

Note that it is not that being old and broken is bad *tout court*. This relativisation of the evaluation of properties to a class might be relevant to help explain non-standard uses of T-terms, such as (118):

(118) I highly value jalopies, because I happen to like old and broken cars[60].

These cases aside, if "jalopy" refers to old and broken cars, and if old and broken cars are bad cars, then having a negative attitude towards the referent of "jalopy" is just normal and natural. It would be just as natural as to have negative attitudes towards the referent of "murderer", or of "shit" (for a similar view developed at length, see Foot 2003).

According this view, the association of an evaluative content to the concept or the term thus has an extra-semantic origin. The possession of a T-term would require no more than knowledge of its truth-conditional content, plus the natural extra-semantic evaluation of the denoted object. It is *because* we know that such and such objects and properties are bad that uses of terms/tokens of concepts referring to these objects and properties trigger negative attitudes or evaluations.

Consider how RBE applies to a few apparently hybrid evaluative terms and concepts. Consider first the concept DOG, which refers to the whole class of dogs. According internalist approaches to meaning and reference determination, it is the descriptive component of DOG that makes it so that it refers to dogs[61]. Building on this basis, some additional descriptive content might be put in a concept or term (here "cur"), so that it will end up targeting only a subclass of dogs that satisfy certain properties.

[60] By contrast, "I highly value murders" is harder, if not impossible, to make acceptable. I thank M. Murez for these remarks.

[61] On externalist views of natural kind terms for instance, it is not the descriptive content of the terms that determines their extension.

The concept of a BIG DOG is one example, as it refers only to the subclass of dogs that are big. Under the view that "cur" are T-terms, the concept of a CUR would just be another such example: it would refer to the subclass of dogs who are dirty, mean, hostile, stray and so on.

Based on such additional descriptive content, the negative attitude associated with the use of such a concept becomes understandable: if curs are dirty and mean dogs, then it is natural that we humans dislike them. Just like DIRT refers to dirt and thus has negative associations for most human beings, CUR refers to e. g. dirty and mean dogs and thus has negative associations.

Similarly, "generous" as a T-term could have a positive flavor because it refers to property of individuals that are inherently likable, or good, and "lewd" could have a negative flavor because it refers to certain bad behaviors or contents.

If seemingly hybrid evaluative terms like "cur" and "boche" aren't co-referential with their alleged "neutral counterpart" (here "dog" and "German" respectively), and if their evaluative content is naturally (extra-semantically) triggered by the evaluation of the denoted objects and properties, then it becomes hard to dissociate the evaluative from the descriptive component.

When a term refers to a property that is objectively bad, obviously the use of the term or deployment of the concept will evoke negative attitudes and evaluations. It will be precisely *because* of the descriptive bit that the attitude is associated, provided that the descriptive bit refers to a property that is already associated with the attitude.

An additional argument in favor of a T-view of slurs like "boche" are what we referred to earlier as "g-contracting" uses of slurs such as (119):

(119) !Angela is German, but she is not a boche.

(119) is non-contradictory only if "German" and "boche" are not coextensional. The existence of such data goes in favor of a T-view of "boche" according to which "boche" targets a subclass of Germans, a subclass which happens to be independently negatively evaluated.

Such inseparability between a term or concept's descriptive content with its apparent evaluative import is precisely what was noted by several authors in the literature on so-called "thick concepts". I shall thus now evoke issues of thickness, before trying to apply RBE to slurs and start considering other views of the additional descriptive element in T-terms.

> Thick concepts hold our interest in part because they seem to unite evaluation and description in some way and, further, make us question what evaluation *is*. But, if that was all that they did, our interest in them would not be as high as it is. They are practical concepts and everyday concepts. They are concepts that pull us - and others - in certain directions and justify some actions and not others. We can use them to shape our world because they seem to be a necessary way of understanding what the world and its people are. If we understand what these concepts are and how they work, we might better understand ourselves and the world we find ourselves in. (Kirchin 2013, p. 18)

There is a class of concepts discussed in meta-ethics, aesthetics and epistemology which raise very similar questions: so-called "thick" concepts. Thick concepts are usually described as "mix[ing] classification and attitude" (Williams 1985, see also Kirchin 2013), which appears strinkingly similar to our earlier characterization of S-terms.

The concept LEWD is a classic example of a thick concept: on the one hand, it refers to a particular objective property (simplifying for clarity, to sexually explicit content or behavior), and at the same time, it is somehow loaded with emotional/moral/evaluative/normative content. In speech, the act of referring to the property (here sexual explicitness) using the term "lewd" - rather than e. g. "sexually explicit" - expresses disapproval, or a negative evaluation towards it.

The same holds with an opposite attitude for the concept GENEROUS. On the one hand it refers to an objective property of an individual (say, the dispositional property to give to the others), and on the other hand, it is weighted with positive attitudes towards that behavior.

Apart from "lewd" and "generous", the following terms have all been viewed as expressing thick concepts: "courageous", "glamorous", "lazy", "triumphant", "selfish", "nasty", "cruel", "truthful", "jejune", "graceful", "lascivious", "lustful", "snitch", "kind", "sublime", "rude", "discretion", "dull-witted", "heroic", "enterprise", "fascinating", "industry", "caution", "idiotic", "assiduity", "grotesque", "kitsch", "clever" "frugality", "observant", "economy", "terrible", "prudence", "reliable", "obscene", "discernment", "promise", "brutality", "coward", "folsky", "lie", "gratitude", "perverted", "mesmerizing", "glorious", "exploiting", "disappointing", "corny" etc.[62] And by exporting the notion of thickness from thought to language, such terms are often called "thick terms".

These emotionally or morally charged concepts were coined "thick" in contrast to "thin" concepts. Thin concepts are one-dimensional, and thus come in two species depending on the nature of the content they involve. SQUARE and TABLE are thin in the sense that they involve only the descriptive dimension. They are *thin descriptive* concepts, because their role in conversation and thought is solely to pick out their referent.

On the other hand, non-cognitivists in ethics believe that concepts like GOOD or WRONG do not pick out any property of the external world, but rather merely express the subject's approval or disapproval for an action. These concepts would thus be *thin evaluative* concepts.

Some take the distinction between thin and thick concepts to be one of degree (Sheffer 1987, Tappolet 2004), others to be one of kind (Williams 1985). Much of the debate evolves around questions of separability[63].

[62] For expository purposes, I will discuss only a few terms ("bitch", "lewd", "cur", "kike", "fag"), but the scope of the conclusions is meant to extend to the whole class of hybrid evaluative terms and concepts.

[63] Traditional noncognitivism take thick concepts to be formed out of two kinds of contents, a descriptive conceptual content and an attitude. Noncognitivists thus claim that moral (or aesthetic) judgments like "Stealing is wrong" just have the appearances of a description of the world, but are in fact of non-propositional form (Blackburn 1984, 1998; van Roojen, 2014). Therefore, truth and falsity is not the relevant dimension for moral judgments (see Richard 2008 for an analogous claim about slurs).

There are different varieties of noncognitivism. *Prescriptivists* equate "Stealing is wrong" with the order "Don't steal!", and equate more generally the attitudinal element in thick concepts with prescriptions or demands (Carnap, 1935, Hare 1952). *Expressivists* argue that "Stealing is wrong" expresses no more proposition than "Stealing, boo!" (Stevenson 1935, The bulk of the debate about thick concepts evolves around what we called "separability". The traditional noncognitivist way of characterizing thick concepts seems committed to what Kirchin 2013 calls "separationism," that is, to the view that one can disentangle thick concepts into their alleged parts.

On the opposite, several nonseparationists authors call that assumption into question based on different observations. For example, Foot (1958a, 1958b) and Murdoch (1956, 1957, 1962) reject the fact/value distinction that they take separationists to presuppose. In doing so, Foot (1958) for instance seems to reach a view similar to RBE, arguing that "rude"

> can only be used where certain descriptions apply (Foot 1958, p. 507),

roughly where the behavior in question "causes offense by indicating lack of respect". According to Foot, if a behavior satisfies the relevant description, then it must simply qualify as "rude". Satisfying a descriptive condition suffices to deserve the name, and no additional negative attitude is needed:

> Refus[ing] to admit that certain behavior was rude because the right psychological state had not been induced, is as odd as to suppose that one might refuse to speak of the world as round because in spite of the good evidence of roundness a feeling of confidence in the proposition had not been produced. (Foot 1958, p. 509)

Blackburn 1993). Note though that there are now expressivist conceptions of propositions for which "Stealing is wrong" would be seen as expressing an expressive proposition (see e. g. Gibbard 2003). *Emotivists* identify the attitude contained in such moral concepts with pure emotions (Ayer 1936).

Another possibility is that these statements, even though they look like a description of an external reality, in fact express another proposition, about the utterer herself. I call this view the *misplaced proposition* view. Under this view, it is not exactly that moral and aesthetic statements are nonpropositional, it is rather that the proposition they express, for some reason, is not the one it seems to be on the surface. For instance, "That is beautiful" could be equated with something along the lines of "I have a positive aesthetic experience towards that".

In other words, "rude" applies as soon as "causing offense by indicating lack of respect" applies. And because "rude" is negatively evaluative, negative moral evaluation applies as soon as the objective descriptive conditions - of causing offense by lack of respect - applies. Foot makes this observation to argue that moral arguments "may always break down"[64]. According to Foot's it is *because* the term "rude" is evaluative that the behavior of "causing offense..." is objectively "bad". The order of determination goes from the evaluativity of "rude" to the moral character of the behavior.

The reference-based evaluation view I just sketched above would rather lean towards the converse order of determination. "rude" would rather evaluative *because* the behavior is objectively "bad". Indeed, according to RBE, "rude" would simply be a descriptive term referring to a certain kind of behavior. As a T-term, it is not an evaluative term *per se*.

It starts being linked to a negative evaluation only after one notices that the kind of behaviors it refers to is extra-semantically evaluated negatively. The term "rude" thus becomes "evaluative" because its referent is "bad", and not the other way around.

Questions about the order of determination - between a term being evaluative and its referent being good/bad - are another debate. What matters to the present discussion is that such an account of T-terms crucially introduces a non-arbitrary, causal link between the evaluative content to descriptive content, similar to non-separationists like Foot who stressed how difficult, if not impossible, it was to keep a neat distinction between a descriptive bit and an evaluative bit in thick concepts[65].

[64] Roughly a form of relativism towards moral statements according to which no evaluative conclusion may rightly follow from purely descriptive premises (see e. g. Stevenson 1944, or Hume).

[65] Note that our current view RBE of T-terms is reminiscent of Vayrynen's pragmatic account of thick concepts (Vayrynen 2013). Vayrynen holds that it is always preferable, whenever it is possible, to postulate that an observed linguistic phenomenon is the result of conversational implicatures, rather than multiplying senses and semantic content (Vayrynen follows Grice 1978 in that sense).

McDowell (1979, 1981, 1987, 1998) reaches similar conclusions in noting that thick concepts are "shapeless". According to McDowell, it is not possible to disentangle thick concepts into parts because one cannot re-characterize the evaluative concept descriptively so that the two concepts are co-extensional. I will briefly come back to issues of paraphrasability when we will evaluate potential objections to RBE in the next section.

Applying that principle to thick concepts, Vayrynen argues that thick concepts are not inherently evaluative at all, but rather derive their evaluative content from pragmatic inferences, "as a function of our communicative and practical interests in discourses involving thick terms and concepts" (Vayrynen 2013, p. *ix*).

Rejecting the thesis that thick concepts are inherently evaluative leads Vayrynen to question the philosophical interest that such terms and concepts have been taken to have, be it in the fact/value distinction debate or with respect to the noncognitivism/cognitivism debate.

A first simple objection can be raised against an absolute version of non-separability, resorting to so-called "objectionable" thick concepts. Many authors take the evaluative content of thick terms to be *defeasible* (e. g. Blackburn (1992), Gibbard (1992), Richard (2008), Vayrynen (2009, 2012, 2013), Eklund (2011)). That is, the evaluative content of thick terms/concepts might disappear - or at least undergo some substantial change - in specific linguistic and pragmatic environments where they are used/deployed. Non-seperability might thus not be absolute, but relative to a context or thinker.

Indeed, it appears that the positive or negative evaluation of thick terms can be perceived as more or less warranted, more or less objectionable, depending on certain factors. Cases like "cur", "jalopy" and "rude" might not have been the best example to start with for that matter, because it is very hard to conceive of a context where the evaluation would not be negative, or of a speaker who would judge the denoted properties to be good.

But it appears that other cases display "evaluative flexibility", as Kirchin (2013) calls the phenomenon. A thick concept displays evaluative flexibility when it can sometimes be used to imply a pro attitude and sometimes be used to imply a con attitude.

I see two possible sources of such flexibility: differences across *contexts* and differences among *speakers*. On the one hand, some features of the utterance context might influence our moral sensitivity and criteria.

On the other hand, there might be certain properties that are typically more prone to intersubjective variation in moral evaluation than others, so that some speakers would typically judge them to be good while others would judge them to be bad. I will now illustrate successively both types of evaluative flexibility.

5.5.1. Contextual Flexibility

Take contextual flexibility first. For instance, Blackburn (1992) remarks that

> one might easily [...] worry that this year's Carnival was not lewd enough (p. 296)

That is, in the context of a Carnival, uses of the term "lewd" do not imply that sexual explicitness is bad (granting that sexual explicitness is an essential characteristic of a good Carnival). Here is another example from Stojanovic (2016), who remarks that

> In the context of movies and works of art, "disturbing", "shocking", and "insane", despite being normally negative, often give rise to positive evaluations. (Stojanovic 2016, p. 3, fn 4)

The evaluation coming with "disturbing", or "lewd", could thus seem sensitive to certain properties of the context (properties that it is not useful to try making explicit here). Such flexibility is a threat to non-separationism because it entails that one and the same term/concept can be used/deployed with different evaluative content, which is conceivable only if the evaluative content is separable from the other - descriptive or reference-fixing - component.

If the descriptive was absolutely *inseparable* from the evaluative, the context could not affect only the evaluative. If we see a case where only the evaluative is affected, then we must conclude that the evaluative is a "component" of the concept, hence that the concept is separable into components.

But these cases are trickier than it might first appear. What is happening in cases of contextual variation such as "lewd" or "disturbing" is not necessarily that the negative evaluation associated with them is traded for a positive evaluation in these contexts (the context of a carnival and the context of a review respectively). Their negative content could well present, but the negative denoted properties could be considered *desirable* in a way.

So in fact, "lewd" and "disturbing" could well systematically convey negative moral evaluation, and the fact that it is a desirable thing to display bad content in the context of a carnival, or to provoke strong emotions in the context of a movie, leads speakers to desire certain things to be more "lewd" or more "disturbing", even though these properties are recognized to be bad to some extent.

These example rather seem to be analogous to utterances such as "I love pain" or even "Pain is pleasurable", where what is meant is not the almost contradictory statement that pain doesn't hurt, but rather that the inherently negative feeling of pain itself can be appreciated on top of its being hurtful, at higher level of evaluation.

We can surely like a bad thing or dislike a good thing, but that does not entail that bad is good or good is bad, as long as we acknowledge a distinction between two levels of evaluation at play in such cases, which seems to be a reasonable assumption. Pain for instance separates into sensory and affective components in certain circumstances, so that some people seem to be able to have positive affective attitudes towards the negative sensory input of pain.

Given this possibility, our examples of contextual flexibility cannot count as evidence for the contextual sensitivity of the evaluative content, and hence not more in favor of separationism.

5.5.2. Intersubjective Flexibility

Now consider a case of intersubjective flexibility. Note first that a speaker might consider the evaluation of a thick term as unwarranted, or objectionable, when it expresses an evaluation that she disagrees with. For example, Gibbard (1992) takes "lewd" as a thick term whose evaluative import is objectionable, because it embodies too prude a view on sexuality for most of us today (the same probably holds for "chaste", or maybe "lascivious").

Similarly, different people could find the negative evaluation conveyed by "lazy" objectionable or acceptable, depending on their views and sensibilities on the purpose of work. Another example could be the French term "navet", which applies to very bad movies. Some people collect and appreciate such movies and could easily say (120):

(120) J'adore les navets.

I love [navets]"[66]

I love rubbishies

But just like above, these are not pure cases of intersubjective variation because the negative evaluation of "lewd" or "lazy" are left unaffected. And in fact, it is even *because* Gibbard acknowledges the negative evaluation of "lewd" that he can judge it to be inadequate or objectionable. The negative evaluation of "lewd" is present even for those who find the evaluation objectionable. Similarly, it is *because* they are *bad* movies that collectors love "navets". We love "navets" in an ironic manner. What we are looking for is rather a case where a single concept is associated with different evaluative contents by different speakers.

"Ambitious" might be close to such a case. Let's assume for the sake of simplicity that the concept AMBITIOUS applies to individuals who have a strong desire of achieving success. That character trait typically receives divergent moral evaluations: some people praise it and others condemn it. Those who find it to be a good character trait associate a positive evaluation with the term and concept, and those who find it negative associate a negative evaluative content to "ambitious".

There is also contextual variation in the case of "ambitious", but it's conceivable that in one and the same context, two different speakers associate opposite evaluative contents to the use of "ambitious".

"Proletarian", or "feminist" might constitute other examples of such thick terms whose evaluative content is highly sensitive to intersubjective flexibility. In a fixed context, two speakers might differ with respect to the moral evaluation they ascribe to any of the properties referred to with these terms.

An aristocrat and a communist will likely tend to associate evaluations of opposite valences to the concept of a PROLETARIAN, just like male chauvinists tend to use "feminist" as an insult, although it conveys a neutral or positive evaluative content for the rest of us. The terms "audacious" or "shrewd" might be other such examples of thick terms sensitive to intersubjective flexibility.

[66] I owe this example to M. Murez.

5.5.3. Back to Separationism

When the property that the concept refers to is susceptible of receiving distinct moral evaluations by distinct groups of subjects, or in different contexts, then it appears the evaluation coming with the concept is separable from its descriptive component. With "rude", it seemed that satisfying a certain descriptive condition was sufficient to deserve the name. From that, Foot concluded that

> there may be the strictest rules of evidence even where an evaluative conclusion is concerned. […] Anyone who uses moral terms at all, whether to assert or deny a moral proposition, must abide by the rules for their use, including the rules about what shall count as evidence for or against the moral judgement concerned. (Foot, 1958, p. 510)

Foot showed that the evaluative "rude" could logically follow from something descriptive like "causing offense by indicating lack of respect", but from that observation she draws the more general conclusion that moral evaluation can follow from descriptive premises, hence that moral arguments can be as valid as arguments about the shape of the earth for instance.

What we now see is that Foot's focus on "rude" might have been misleading. Because "rude" refers to a behavior causing offense by indicating lack of respect, it is almost unanimously judged as bad. Hence, inference from a descriptive statement O (involving the descriptive paraphrase "causing offense by indicating lack of respect") to an evaluative statement R (involving the thick term "rude") seemed to support an inference from "causing offense…" to "bad".

Similarly, since being a stray, mean and dirty dog is unanimously considered to be bad in virtually every context, and since CUR happens to refer to stray, mean and dirty dogs, then the concept seemed to be inseparable from its negative moral evaluation.

Unlike for "rude", we see that satisfying a descriptive condition does not actually suffice to deserve the name or "lewd" or "ambitious". Indeed, if one were to try reconstructing Foot's argument with "lewd" or "ambitious", although R would follow from O ("sexually explicit" would entail "lewd", and "strongly desirous of achieving success" would entail "ambitious"), it would not follow that sexual explicitness or strong desire of success is good, nor that it is bad. Maybe such inferences are valid only with respect to a certain fixed context?

Such contextual and intersubjective relativity in the evaluative content associated with certain concepts show that the evaluation does not directly follow from the satisfaction of the property described in the concept. There are properties whose moral evaluation is less clear, and varies among subjects (e. g. strongly desiring success appears to be bad to some, and neutral or good to others), or among contexts (e. g. sexual explicitness is less inappropriate in the context of a carnival than in everyday life).

This relativity of moral evaluation could lead us to build a level of context-sensitivity into the semantics of these terms and concepts (e. g. contextualism or relativism), but that is not the object of the present discussion.

Here, the fact that moral evaluation is relative to speakers and contexts is simply taken to allow us dissociating more clearly the term's descriptive content from its evaluative import. The two elements seemed inseparable because of the focus on T-terms referring to properties whose evaluation is of common agreement, but uses of T-terms displaying evaluative disagreement across speakers and context show that they aren't necessarily inseparable.

5.6. Two Unsuccessful Objections to a Reference-Based View of Evaluation

5.6.1. Back to Projection

Objection

The dissociation between the descriptive and the evaluative components of T-terms helps us point to a potentially incorrect prediction of RBE. Because RBE claims that thick terms have their negative import naturally associated with the subclass they refer to, it predicts that T-terms and their descriptive counterparts equally project.

If "rude" has no *conventional* meaning over and above that of "causing offense by indicating lack of respect", and has its evaluative content derived pragmatically from one's evaluation of the property itself, then "rude" and "causing offense by indicating lack of respect" are just synonymous in both the descriptive and the evaluative aspects, and should thus behave alike in all linguistic environments.

This appears to be a wrong prediction of RBE. Consider first:

(121) a. John's remark was not rude.

→ If it were it would have been a bad thing.

b. John's remark did not cause offense by indicating lack of respect.

↛ If it had it would have been a bad thing.

Testing projection, we focus on the inference that the speaker judges the kind of behavior considered to be bad. It seems that (121a) and (121b) display a contrast with respect to that inference. A speaker uttering (121a) seems to be committed to a negative evaluation of causing offense by indicating lack of respect.

Were John's remarks to fall under the extension of "rude", it would be a bad thing on the part of John. That inference is not triggered by a use of (121b). The speaker could well continue her utterance with the following qualification in (122b) without giving rise to oddity in, whereas that is not the case of (122a):

(122) a. ?John's remark was not rude, although there would be nothing bad whatsoever if it had been.

b. John's remark did not cause offense by indicating lack of respect, although there would be nothing wrong whatsoever if it had.

Consider also:

(123) a. *Lolita* is not lewd.

→ If it was it would have been a bad thing.

b. *Lolita* is not sexually explicit.

↛ If it was it would have been a bad thing

Maybe even more than in the previous pair because "lewd" is more objectionable than "rude", there is a contrast in projection. (123a) conveys a negative evaluation of sexual explicitness and (123b) does not.

If the evaluative content of "lewd" came from the (usual) negative evaluation of the property it refers to as RBE describes it, then referring to the very same property with a phrase other than "lewd" should trigger the same effects. But contrary to (123a), one cannot infer from the use of (123b) that the speaker has prudish views on sexuality.

Given such contrasts in the projection of evaluation between T-terms and co-extensional paraphrases thereof, the claim that evaluation starts with a rich reference-fixing bit in the concept is threatened. The evaluative content of hybrid evaluative concepts cannot originate solely in the evaluation of the property, for this fails to account for the projection facts. Initially, it is projection that motivated a theoretical comparison of slurs with thick terms, but it seems that data as simple as (121)-(122)-(123) suffice to contrast the two projection profiles.

There are two ways to address this worry. First, nothing suggests in principle that T concepts should be paraphrasable. Just because their encoded content is purely truth-conditional/reference-fixing does not mean that one must be able to construct a co-extensional paraphrase of it, or even a co-intensional paraphrase for that matter.

Indeed, even everyday (thin) descriptive concepts like TABLE are notoriously hard to paraphrase. Consequently, the contrasts displayed in (121), (122), and (123) could simply stem from an oversimplification of the paraphrase that is tested.

Another similar although less radical answer to the worry is to note, not that concepts are unparaphrasable, but at least that the paraphrases are expected to be very fine-grained and complex in order to be truly co-extensional with the relevant T-term. Unless we compare the projection behavior of the tested T-term with that exact complex paraphrase, no apparent contrast should be taken to constitute evidence against RBE.

Indeed, it was even clear from the start that "causing offense by indicating lack of respect", or "sexually explicit" were mere approximations for respectively "rude" and "lewd". The real life uses of "rude" and "lewd" apply to way more diverse and complex sets of behaviors, contents, events and so on.

So, if we are unable to find a suitable paraphrase to test, is there a way to test the predictions of RBE as to projection? There might be a way to do so, with the help of a demonstrative phrase instead of a (not so-)co-extensional paraphrase.

Imagine a situation where Sue made a stereotypically "rude" remark to her interlocutor in an earlier conversation, in the presence of two bystanders. She might have for instance overtly and inconsiderately criticized her interlocutor's physical appearance.

Imagine now that in a later conversation, John made a comment on his interlocutor’s appearance, in the presence of the same two bystanders. These two bystanders are now discussing the status of John's comment, manifestly comparing it to the still salient rude remark Sue made earlier (124b):

(124) a. John's remark was not rude.

b. John's remark was not like *that*.

Assume it is manifest in the context of the conversation that the demonstrative refers to the behavior that Sue displayed earlier. In that case, the demonstrative could be said to be co-extensional with "rude", without actually using the term expressing that concept. Comparing projection in the two cases might then be a way to test the prediction of RBE.

Because (124a) is the same as (121a), the inference to the effect that the speaker believes it is bad thing to have such behavior still projects. But this time, unlike (121b), the same inference seems to project in (124b): were John's behavior like "that", it would be a bad thing.

The same thing can be done for "lewd". Imagine again that both speakers agree that a certain movie is stereotypically "lewd", and now discuss the novel *Lolita* comparing it to that movie:

(125) a. Lolita is not lewd.

b. Lolita is not like *that*.

Here, like in (122), although linguistic intuitions are admittedly quite hard to access, it is not so clear anymore that the two utterances display different projective inferences. The contrast presented in (121), (122) and (123) are therefore a weaker threat to RBE as one might have first thought.

Overall, it is conceivable that a finer-grained paraphrase ends up projecting the evaluative content associated with the tested T-term. And even if the judgments happened to be unclear for better paraphrases (like with the demonstrative), these contrasts would still be insufficient against RBE in general, because T-terms and concepts could unproblematically happen to be unparaphrasable.

5.6.2. Co-Extensional Concepts with Different Valences

Objection

Let us now investigate a second straightforward objection to RBE. Start by considering again "generous". We just saw that according to RBE, "generous" would have a positive evaluative content because it refers to the property of an individual (say: being disposed to give to others) that is inherently likable.

That being said, take now the concept PRODIGAL. "Prodigal" seems to apply to the same property of an individual as "generous", but it comes with negative evaluation. Where "Prodigal" conveys a negative evaluation of the denoted behavior, "generous" is loaded with positive evaluation. So doesn't the pair "generous"/"prodigal" constitute a counterexample to RBE?

Indeed, how could it be that the evaluative content of a T-term is reference-based if there are co-extensional concepts with evaluative contents of opposite valences? If the evaluative content originated in the evaluation of the property itself, as suggested by RBE, there should be no such variation in valence.

For a certain evaluative content to be systematically associated with a term or concept referring to a property, that property must either be good or be bad. And if the property of being disposed to give a lot to others was judged to be positive by some speakers in some contexts, and negative in other contexts by other speakers, then RBE would be resourceless to explain why "generous" always comes with a positive evaluation and "prodigal" with a negative evaluation.

RBE would even make the false prediction that speakers who find the denoted behavior to be bad in the context could use "generous" to refer to it and convey a negative evaluation, and respectively that the speakers who find the property of giving a lot to others to be good in the context could use "prodigal" to refer to it in a positively tinted manner. This prediction seems outright false. One cannot in fact use "generous" negatively and "prodigal" positively.

Or at least, even if one can find many utterances involving a negative evaluation of generosity such as "Generosity is a bad thing" or "Being generous is weak and foolish", it is not clear at all how exactly the negative evaluation of the denoted behavior comes across, as we discussed above.

It could well be that "generosity" conveys a positive judgment that is then cancelled at a higher level by the predicate "is a bad thing". This does not entail that "generosity" is used with a negative evaluative content.

So it looks like RBE might only be able to explain the evaluative content of T-terms referring to properties whose evaluation is to some extent shared across speakers and across contexts. If the evaluation of a property is sensitive to contextual and intersubjective variation, then the evaluative content of a concept referring to that property must have another source than evaluation of the property itself.

Reply to the Objection

One way to answer this worry is to maintain that, appearances notwithstanding, the two concepts aren't in fact co-extensional. We could argue for instance that among the class of individuals with a propensity to give a lot to others, "prodigal" in fact applies only to those who give *too much*, and "generous" to the others.

Although it might be the case that one and the same behavior can be judged good by some speakers and bad by others, or even good by a speaker in a context and bad in another context by the same speaker, it is not the case that a single speaker in a single context can equally apply "prodigal" and "generous" to the behavior.

Once a perspective with a subject and a context are fixed, "prodigal" and "generous" have different extensions: they split the relevant class of all giving-behaviors in two mutually exclusive subclasses. With different extensions and different extra-semantic evaluation of the respective referents, the difference in evaluative import of these terms follows directly: as giving to others is usually perceived as a good thing - unless it reaches a certain threshold after which it is considered to be a bad thing, "generous" and "prodigal" will be tinted with respectively positive and negative evaluative contents[67].

[67] A way out of this objection could be to build context-sensitivity into the semantics of these terms and concepts, as I suggested earlier. Taking back the example of "prodigal" applying

5.7. Slurring Concepts Aren't T-concepts

5.7.1. Ambiguity

As we saw, slurs are usually described as terms targeting groups or individuals on the basis of gender, ethnicity, sexual orientation and the like, and they are of interest because - at least pre-theoretically - they seem to fulfill two roles at once: picking out a referent (or at least being *about* certain individuals), and expressing, or signaling, certain attitudes, or implicit beliefs, or emotions on the part of the speaker.

The terms "boche" and "German" are about the same individuals in the world, but the former is depreciative and offensive in a way that the later isn't. Equivalently, the concepts BOCHE and GERMAN seemed to pick out the same individuals in the world, but the former involves an evaluative content that the latter lacks. One of our goals was to understand how the two dimensions combine, in speech and in thought.

Given the above discussion, a reduction of slurring concepts to thick concepts understood as-T-concepts is *prima facie* tempting[68]. Slurs and thick terms indeed seem to be of interest to the philosophers and linguists for roughly the same reason: the combination of descriptive and evaluative contents.

to those who give "too much" whereas generous apply to the rest of people who give. We shall note that given that what counts as too much must be evaluator-sensitive, saying this might in fact be compatible with coextensionality: it may be the very same individuals who, according to some, give just enough and, according to others, give too much.

[68] Note that wondering whether slurs express T-concepts, or are special uses of thick concepts, is different from, although not orthogonal to, the question of their respective analysis. If slurs do express thick concepts, then surely they should receive the same treatment. But if slurs and thick terms show significant distinct properties and linguistic behaviors, the way is open to different structural analyses.

But I will now argue, based on Nunberg's observation that many slurring terms are ambiguous, that RBE cannot be suitably applied to slurring concepts.

So could slurs be T-terms? A positive answer seems to be suggested by g-contracting uses of slurs such as (119). There exist uses of slurs as clear T-terms, where we see that the stereotypical additional descriptive element seems to play a reference-fixing role. As Camp puts it, in such g-contracting uses, "the slur's extension is restricted to stereotype conforming members" (Camp 2011).

A typical example is Chris Rock's outrageous but non-contradictory "I love black people but I hate niggers". For it to be non-contradictory, "black people" and "niggers" must have different extensions, and it seems that we can retrieve a reading where the term "niggers" applies to a subgroup of black people having certain additional traits.

Take another standard example of slurs such as "kike" and "boche", and try to apply RBE to account for the evaluation they appear to carry. If "boche" was a T-term, it would roughly contain the three following bits: (i) a descriptive content equivalent to that of "German", (ii) an additional descriptive - reference-fixing - element motivating (iii) a (independent) negative evaluation of the target.

So according to the first two tenets of RBE, slurs are not co-extensional with their Neutral counterparts (NCs), and this seems confirmed by g-dontracting uses of slurs. In particular, "kike" and "Jew" aren't co-extensional, and neither are "boche" and "German".

According to the second clause of RBE, "boche" is supposed to refer to a subclass of German people with certain bad properties, and "kike" to a subclass of Jewish people with certain bad properties. What could these properties be? A likely possibility is that these additional properties correspond to the negative stereotypical properties that slurs users take their targets to have.

For instance, if there is a Germanophobic stereotype that German people are cruel (see e. g. Dummett 1973), we could specify RBE applied to "boche" as follows: where "German" refers to the whole group of German people, "boche" has a richer descriptive content somehow involving cruelty (clause (i)).

The way in which cruelty is involved, according to RBE, is that it makes it referring only to the subgroup of cruel German people (clause (ii)). And given that cruelty is unanimously taken to be a bad property to have for an individual, a certain negative evaluative content will be naturally (and extra-semantically) associated with uses of the slur "boche" (clause (iii))[69].

But there is a sense in which RBE is not fully satisfactory when applied to slurs, and this is not because of the way in which it derives their evaluative component, but rather because of the nature of the rich descriptive component it ascribes to them. It is thus the second, and not the third clause of RBE that seems unsatisfactory, as one shall now see.

In fact, the view RBE has about the reference of slurring concepts does not do justice to the strong intuition that "kike" is used as a slur against all Jewish people, and "boche" against all Germans. "Kike" is an anti-Semitic insult, not an insult towards only specifically dreadful Jewish people. Nunberg (2016), for instance, remarks that many slurs appear to be ambiguous:

> When Lil Abner says, 'I went out with a lot of bitches', his utterance is ambiguous: he might mean either that he dated a lot of nasty or unpleasant women, or just that he dated a lot of women. In the first instance you can contest his utterance by rejecting either component; you can say either 'That's not true; you've never been out on a date' or 'No, your dates were always considerate and good-natured.' But if he's using *bitch* simply as a derogative for women in general you can only make the first objection; you can't say 'Well, it's true you went out with a lot of women but they were all very nice.' (Nunberg 2016, appendix p. 61)

Here, Nunberg acknowledges that there is a reading of "bitch" which works in the way RBE predicts: it applies to a subclass of women (who are "nasty or unpleasant") and is thus extra-semantically associated with negative evaluation because unpleasantness or nastiness are bad properties.

[69] Note that there is a potential issue in that "cruel" might be an evaluative notion itself, but we can safely put that issue aside and consider instead "cruel*'", an hypothetical neutral descriptive counterpart of the potentially evaluative "cruel".

But Nunberg notices that there is another reading of "bitch", still evaluative, which does not work in this manner. On this second reading of Lil Abner's utterance, the use of "bitch" applies to women in general, and is thus co-extensive with the counterparT-term "women".

It thus appears that "bitch" has in fact not one, but *two evaluative* readings. There is a reading I will call *narrow,* on which it targets only a subset of women, and a reading I will call *wide,* on which the term targets women in general. RBE could easily account for the narrow reading of "bitch", but it is inapt to account for its wide reading.

If "bitch" has a wide reading under which it refers to women in general, there is no way one can derive its evaluative content as is done by the third clause of RBE, because there would be no additional property denoted by the wide use to motivate its negative evaluative content.

And if two such readings of T-terms were systematically available, it might suggest that there are in fact at least two types of evaluative terms, which goes against an assimilation of slurring concepts to T-concepts as described in RBE.

5.7.2. Redundancy and Contradiction

If there really are two kinds of evaluative concepts, those with a rich reference-fixing component (T-concepts), whose evaluative import stems from a *narrow* extension plus a natural evaluation of their referent, and those which are co-extensional with their counterparts (S-terms), whose evaluative content is added on top of their *wide* extension, then we expect to be able to construct a test distinguishing between the two.

Nunberg (2016) alludes to such test in noticing the following contrastive judgments, which he takes as evidence that the evaluative component of certain terms (that he calls "appraisives", our "narrow" uses) is conventional, unlike that of (wide uses of) slurs:

> As Sadock (Sadock 1978) notes: 'Since conversational implicatures are not part of the conventional import of utterances, it should be possible to make them explicit without being guilty of redundancy'. In this regard prejudicials contrast with the

> appraisive words with which they're often lumped, where the evaluation is genuinely part of the word's meaning and hence can't be nonredundantly predicated of it. Utterances like 'Toadies are obsequious', 'Fleecing someone is unfair' and Shrill sounds are unpleasant' are likely to elicit the reaction 'So what else is new?'. (Nunberg 2016, pp. 14-15)

Elaborating on this remark, Nunberg's test consists in constructing an utterance with an hybrid evaluative term followed by a paraphrase of its evaluative content, and to observe speaker's intuitions of redundancy. In my terminology, it will be a T-term (with a narrow extension) if speakers find the resulting utterance redundant, and a S-term (with a wide extension) otherwise. Consider first:

(126) !These bitches are nasty

As there are two readings of "bitches", one wide and one narrow, (126) is ambiguous between a narrow and a wide reading. Note that (126) will typically be perceived as redundant under the narrow reading, whereas it sounds felicitous (the attitude aside, of course) under the wide reading. This confirms the distinction between narrow and wide uses of T-terms, or the distinction between "general pejoratives" and "slurs" which can be found in Hay 2013.

Compare now:

(127) Toadies are obsequious.

(128) !Kikes are greedy.

Most speakers will treat (127) as trivial or redundant, whereas it does not seem to be the case of (128), as uttered by an anti-Semite.

The sense of redundancy created by (127) could be taken as evidence that the property of being obsequious is somehow already encoded in the concept TOADY, and conversely the felicity of (128) would indicate that the property of being greedy is not strictly speaking entailed by the concept KIKE, but is rather pragmatically connected.

Such a contrast indicates a potentially important structural difference between T-terms and concepts on the one hand, and Slurring terms and concepts on the other hand. And since redundancy is just the other side of contradiction, we can construct a twin test to tease T-terms apart from slurring terms. Nunberg adds:

> In the hip hop [narrow] sense of the word, there's no contradiction in saying "That bitch is kind and sweet" though that utterance sounds contradictory if *bitch* is being used as a routine pejorative [targeting women]. (Nunberg 2016, p. 61)

So just like certain evaluative terms predicated of their explicit evaluative content creates redundancy, utterances of these terms followed by the explicit negation of their evaluative content creates a sense of contradiction. These are two sides of the same coin.

The problem for RBE that I raised here is this: RBE might work for the narrow T-terms, but it is resourceless to account for slurs. Indeed, RBE crucially connects the evaluation to a rich reference-fixing component that is not identical to the concept's counterpart.

If there are (readings of) T-terms that still refer to the whole class and are co-extensional with their counterparts, then there is no additional reference-fixing component anymore to motivate evaluation. Something else is needed then to account for slurring terms, which are evaluative although co-extensional with their counterparts.

In a nutshell, it looks like slurring terms cannot be T-terms because their additional descriptive element, the one that motivates the evaluative character, does not elicit redundancy judgments, and is thus *not* reference-fixing. Slurs are on the opposite coextentional with their counterparts and target the whole supergroup, even though it is loaded with more descriptive content than their counterparts. There must therefore be a sort of descriptive though not-reference-fixing content at play in slurs.

I conclude from these observations (ambiguity and redundancy/contradiction) that g-contracting uses of slurring terms, where it is clear that slurs are used as T-terms, are marginal and shall thus not be taken on board too hastily. They do not so much argue in favor of a T-view of slurs than for a T-view of g-contracting uses of slurs. The central cases do not seem to conform to the T-model.

Given that RBE ended-up not being fully satisfactory for slurring terms, I will now investigate another view of slurring terms and concepts according to which all dimensions of slurs are put in their truth-conditions. According to RBE, the descriptive content of SCs and STs is rich, but it is not evaluative *per se*. The evaluation is still derived externally; it is somehow a consequence of the descriptive meaning when combined with our moral faculty. The next view will try to locate even the evaluative dimension into the concept's content. Let us now turn to the evaluation of such a radical truth-conditional account of slurring concept.

Chapter 6. A Radical Non-Hybrid Account of Slurring Concepts

In the present chapter[70], I consider a first radical version of a non-hybrid account of Slurring Concepts. The goal of this chapter is to provide arguments based on linguistic evidence that discard a truth-conditional analysis of STs and pave the way for more promising approaches.

To this effect, I focus on Hom's (2008, 2010, 2012) and Hom and May's (2013, 2014, 2015) view that slurs express complex, though one-dimensional, predicates. Although critics of the truth-conditional account of STs (TCA) typically target the arguments based on "substitutability data", where it seems that STs cannot be substituted *salva veritate* with their NCs, they overlook the important set of what I call "non-projectability" data, erroneously taken by TCA as evidence that the evaluative content of STs does fall under the scope of truth-conditional operators.

I aim at bridging this gap. I present TCA and consider Hom and May's analogy between STs and fictional terms before discussing TCA's neglect of projection facts. This leads to the positive contribution of this chapter: I present novel evidence showing that non-offensive uses of embedded STs are the result of metalinguistic effects, and put forward new data involving Absurd Counterfactual Conditionals (ACCs).

Eventually, I discuss Hom and May's distinction between *offense* and *derogation*, aimed at explaining projection, and bring about the example of STs for fictional entities to argue that this move fails to reach its goal.

[70] This chapter is based on joint work with Cepollaro.

Summing up Frege's hybrid view on evaluative concepts, on the one hand there would be a logically relevant semantic content of "rose" and "cur" that both the botanist and the poet access to (sense), and on the other hand there would be their respective subjective "coloring" (be it private or public) of that semantic content which is making no logically relevant contribution (tone).

We saw that Frege's view on concepts like CUR could be seen as the ancestor of modern hybrid-expressivists *à la* Potts. The term "dog", having the same sense but no color/tone would be the neutral counterpart of the term "cur".

The newly developed semantic models of hybrid-expressivists aim at systematically deriving "tone" in cases where it is conventional (as it is with "cur" and "but"), and therefore build on Frege's primary intuition, as well as on Grice's notion of conventional implicature.

Now, the earlier objection I raised earlier against hybrid accounts of STs can be summed up as follows. The hybrid accounts posit two mental ingredients (a descriptive concept, and an evaluative attitude), which combine in virtue of the conventions of language. This makes the phenomenon an essentially linguistic phenomenon.

But if we want to make room for the possibility that the phenonenon might not be purely linguistic, but also mental, we must either find another way of combining the two ingredients without relying on linguistic conventions, or we must give up the hybrid approach and look for another type of account.

Another type of account has, indeed, been put forward in the literature. Remember Dummett's inferential account, with the introduction rule "German —> Boche" and the elimination rule "Boche —> Cruel" etc.

A non-hybrid account would take a ST like "boche" to express a descriptive concept along the lines of "cruel etc. because German". An individual *x* satisfies the concept if and only if *x* *i)* is German, *ii)* is cruel etc., and *iii)* there is an intrinsic connection between the two properties: one derives from the other.

Hom (2008, 2010, 2012) and Hom and May (2013, 2014, 2015) have elaborated such a radical non-hybrid account, and emphasized a consequence that they endorse: such a concept is bound to have an *empty extension*, because the third condition can't be satisfied: there is no intrinsic connection between nationality or ethnicity and morally objectionable properties like cruelty etc.

Hom and May happily endorse this consequence, but there is a significant drawback: we have to give up the view that STs like "boche" often have a NC with the same extension ("German"). Since the extension of STs is always empty, on the sort of view I have just summarized, it follows that slurring concepts do not have a NC either. As we shall see now, the view put forward by Hom and May faces additional problems. In particular it seems that it cannot account for projection facts.

First consider the following view:

> **Truth-conditional-stereotypical-evaluative view (TCSE):** STs express concepts with a rich reference-fixing component, including a stereotype-predicate and an evaluative operator.

Applied to a slurring concept like KIKE, TCSE is not very different from Hom (2008, 2010, 2012)'s view who gives pride of place to the stereotype:

> For example, the epithet "chink" expresses a complex, socially constructed property like: ought to be subject to higher college admissions standards, and ought to be subject to exclusion from advancement to managerial positions, and …, because of being slanty-eyed, and devious, and good-at-laundering, and …, all because of being Chinese. (Hom 2008, p. 431)

Similarly, "kike" would not refer to the whole class of Jewish people, but instead, only to a potentially empty subclass of Jewish individuals who would happen to satisfy the rich and complex set of stereotypes.

As soon as we have a richer descriptive content, and that the richer descriptive content includes an evaluative operator such as "ought to be the target of negative moral evaluation", "worthy of contempt", "despicable" etc., I remark that the stereotype in itself does not do much work. TCSE can thus be simplified as follows:

> **Truth-conditional-evaluative view (TCE)**: STs express concepts with a rich reference-fixing component, including a standard predicate and an evaluative operator.

TCE is roughly the view defended by Hom and May (2013, 2014, forthcoming), according to which STs express complex (pejorative) predicates constructed out of a second order pejorative predicate (PEJ) taking a first order standard group-referencing predicate as its complement. In what follows, I shall focus on this sort of a truth-conditional account of SCs.

The aim of any truth-conditional account of slurring concepts (TCSE or TCE, henceforth TCA) is to locate the *evaluative content* at the level of the reference-fixing component. According to TCA, the pejorativeness of SCs lies at the truth-conditional level: for instance, the SC BOCHE would make the same truth-conditional contribution as a complex evaluative predicate "German and worthy of negative moral evaluation because of that", or something along these lines.

SCs and this kind of paraphrases would thus be synonymous[71], so that the one could be substituted with the other in sentences/thoughts they appear in without affecting their truth-conditions[72].

TCA can therefore be described as a reductionist theory of slurirng concepts: it aims at reducing their "hotness" to the descriptive level without calling on other mechanisms or dimensions of meaning. For TCSE, TCE and other versions of TCA, SCs are negatively loaded simply because they straightforwardly ascribe negative moral properties to individuals, and so are STs.

So Hom and May (2013, 2014, forthcoming) defend a version of such a view: they paraphrase KIKE with something along the lines of "ought to be the target of negative moral

[71] I will simply equate *synonymy* between *a* and *b* with identity of their intensions (or Kaplanian *contents*): *a* and *b* map the same worlds to the same individuals, they have the same descriptive material. Coextensivity follows (although there can be coextensivity without synonymy, like "Barack Obama's predecessor" and "Bill Clinton's successor" for instance).

[72] Hence the discussion on "substitutability data" found in Copp and Sennet 2015.

evaluation because of being Jewish", so that the concept KIKE entails its neutral counterpart JEWISH.

Hom and May posit the existence of a function turning neutral concepts into their slurring counterparts with the help of a silent operator: PEJ. Predicates like "is a kike", "is a dirty/fucking/damn Jew", or even "is Jewish" accompanied by an expression of disgust, are all viewed as different externalizations of one and the same function, PEJ(Jew), meaning something like "ought to be the target of negative moral evaluation because of being Jewish[73]". That is what I will call the semantic claim:

> **Semantic claim**: The pejorative content of SCs is part of their truth-conditional content (e. g. WOP = PEJ(ITALIAN[74])).

We can here distinguish between three components in the complex predicates SCs are supposedly made up of, under such a view: i) a reference to the target class, ii) a moral negative evaluation, and iii) a (causal) connection between the two. Take (129) and (130) for instance:

(129) !Leonardo da Vinci was a wop.

(130) !Leonardo da Vinci was a limey.

According to this version of TCA, (129) and (130) are false in all worlds and all contexts, for no one is worthy of negative moral evaluation because of belonging to a group (be it Italians or British people). Now, it shall be noted that even though (129) and (130) have the same

[73] Note in passing that this solution under-generates recursive uses of expressives. It indeed predicts the following extravagant meaning for complex expressions like "dirty kike": "ought to be the target of negative moral evaluation because of being ought to be the target of negative moral evaluation because of being Jewish" - at least without *ad hoc* restrictions on PEJ.

[74] "PEJ(ξ) functionally combines with any characteristic counterpart term, t, typically designating race, gender, religion, class, and so forth, to form a pejorative, PEJ(t)" (Hom and May, 2012, p. 6)

truth-value, an utterance of (130) is misleading in a way that an utterance (129) is not. Hom and May's 2015 version, to which I now turn, explains the contrast.

As it stands, TCA is merely making a *semantic claim* that ascribes a certain lexical entry to certain expressions. Hom and May pair it with an additional *moral claim*:[75]

> **Moral claim**: necessarily, no one ought to be the target of negative moral evaluation because of belonging to a group.[76]

It follows that SCs have a necessarily empty extension, and this makes STs and SCs similar to fictional terms and concepts (Hom and May 2015). As just noticed, there is a sense in which (129) is less misleading than (130), although they are both false.

The analogy Hom and May draw between slurs and fiction accounts for this contrast: there is fictional truth on the one hand, like "Unicorns are white", and material truth on the other hand, like "Horses are four-legged": utterances like "unicorns are white" are *fictionally* true but *materially* false (or at least *not true*).

In the very same way, Hom and May take (129) to be *fictionally* true (in the fiction of Italianophoby) and *materially* false; whereas (130) is both *fictionally* and *materially* false. According to the authors, the intuition that "Leonardo is a wop" is (strictly speaking) true "embeds a mistake of fiction for fact" (Hom and May 2015).

To illustrate the parallel, Hom and May consider the following case: in the Middle Ages, people took tusks of narwhals to be unicorn horns, to which they ascribed magical properties (neutralize poisons, heal diseases, etc.). In other words, unicorn-believers took some real

[75] The authors discussed do not explicitly make this distinction, which is a reconstruction of their view.

[76] B. Bantegnie warned me that the moral claim could be false simply because we can put bad actions in the criterion of individuation for the group (e. g. the group of serial killers). Even in such a case though, if a member of such group ought to be the target of negative moral evaluation, it is because she is a serial killer, and not because she belongs to the group of serial killers.

objects - narwhal tusks - to be unicorn horns, by ascribing them properties that the real objects didn't have.

The same goes for slurring concepts: anti-Semites mistakenly take Jewish people to be KIKES, by ascribing to them properties that they lack as a matter of necessity (like being contemptible because of being Jewish). That is to say, for Hom and May, anti-Semites are fiction-believers.

Let us now turn to the general difficulties that such a radically truth-conditional account of slurring concepts face, especially in explaining the projection behavior that uses of these concepts display when used in language.

6.2. The Problem of Projection

In this section, I assess the main problem that TCA meets in accounting for the offensive content that expressions of slurring concepts convey in embedded environments. Consider the following pairs:

(131) a. John is a faggot.

b. John ought to be the target of negative moral evaluation because of being homosexual.

(132) a. !John is not a faggot.

b. John ought not to be the target of negative moral evaluation because of being homosexual.

As we saw in the previous chapters, although (131a) and (131b) might seem to be synonymous at a first glance, they display crucially distinct behaviors under their negated alternatives (132a) and (132b) respectively. Whatever expressive content the ST "faggot" conveys, it still conveys it under negation in (132a), whereas (132b) is not derogatory. There is therefore something derogatory in the content of "faggot" which is *not* affected by negation: it *projects*.

STs have a very broad projection profile, as we saw: their pejorative content scopes out of most (if not all) truth-conditional operators, from modals (133) to conditionalization (134), questions (135), quantification over events (136) and so on.

(133) !John's girlfriend could very well be a wop.

(134) !If Mary is a dyke, then she won't like that dress.

(135) !Is Bob a faggot?

(136) !Every time I meet three chinks in a row, I suppose I'm in Chinatown.

Moreover, according to TCA, "John is a boche" is always false. From this - plus the standard analysis of negation - it follows that "John is *not* a boche" is always true. Similarly, each of the following utterances is necessarily true according to TCA:

(137) !Obama is not a chink.

(138) !There are no kikes at my office.

(139) !If Freddie Mercury was a fag, then he was worthy of contempt.

For TCA, derogation consists in the ascription of a negative moral property to a subject. In (133)-(139), speakers detect a pejorative content even if no negative moral property is *ascribed* to subjects (i. e. even if the expression is not strictly speaking *predicated* of the subject). Not only are (133)-(139) predicted by TCA to be true, they are predicted to be non-derogatory.

To many competent speakers, (137)-(139) for instance, sound as derogatory as any utterance of "Obama is a chink", "There are kikes at my office" or "Freddie Mercury was a fag". TCA is not well-equipped to account for these projection facts.

To sum up, if the evaluative content of SCs is reducible to negative-property-ascription - as TCA, TCSE and TCE claim - then there should be no derogation in utterances of STs that do not involve negative-property-ascription. But speakers *do* detect negative evaluative content in such cases, therefore the evaluative content of SCs and STs cannot be the result of predication only.

The debate therefore crystallizes on utterances of STs under negation and other operators: TCA predicts that such utterances have a non-derogatory reading, but speakers judge them derogatory. The dispute seems to be about what we mean when we judge an utterance to be "derogatory". If "derogatory" means, as Hom and May define it, the ascription of a negative moral property to a group or individual, then surely "John is not a boche" is not "derogatory" in that sense, because "being a boche" is not *ascribed* to John.

But redefining the folk notion of "derogation" as the ascription of negative moral properties does not seem to capture the phenomenon of slurring and pejorativeness we are after. I come back to this point later.

Hom (2008) and Hom and May (2013, forthcoming) provide two answers to the projection problem. On the one hand, they deny the intuitions that anything projects out of e. g. negation in "John is not a kike", and rely on non-pejorative readings of embedded slurs like "There are no kikes" to argue that projection is an illusion, that negation in these cases in fact successfully took scope over the evaluative content of STs. I will show by means of three linguistic tests that such readings are in fact brought about by an intervening metalinguistic factor.

On the other hand, Hom and May acknowledge that utterances of STs under e. g. negation still have a certain offensive flavor. They put forward a distinction between "derogation" and "offense" to explain the projection intuitions. I put forward new data involving new examples of STs for fictional entities to argue that this move fails to reach its goal. In a nutshell, my main arguments against TCSE, TCE and TCA in general are thus the following:

> (i) Data involving non-offensive occurrences of embedded STs are systematically metalinguistic - which explains the confusion on "non-projectability" data.
>
> (ii) TCA makes wrong predictions about the behavior of STs, in particular under Absurd Counterfactual Conditionals (ACCs).
>
> (iii) The attempted distinction between derogation and offense is ineffective, as shown by the existence of STs for fictional entities.

I acknowledge that some occurrences of embedded STs such as "There are no kikes", have a non-pejorative reading available. Nevertheless, I argue that such readings are not supporting TCA, as they are actually brought about by metalinguistic effects (or other perspectival effects[77]).

Consider the following statement:

(140) Yao Ming is not married.

There are (at least) two readings of (140), depending on how negation is interpreted. A *propositional* interpretation of negation gives rise to a reading under which Yao Ming has no husband or wife; whereas a *metalinguistic* interpretation of negation brings about a reading under which the speaker finds that the predicate "married" isn't accurate to describe Yao Ming's marital status.

The two readings can be made explicit with the following continuations:

(141) a. Yao Ming is not married, he is single. (PROPOSITIONAL NEGATION)

b. Yao Ming is not married, he is joined in holy matrimony. (METALINGUISTIC NEG.)

[77] Jeshion (2013a) makes a similar observation. Note that the purpose of this section is to investigate whether *standard* embedded uses of STs can be non-pejorative; a separate issue that I set aside is whether some STs can have non-pejorative uses *in general*. An interesting case is the case of appropriation, where members of a targeted group use the ST with other in-groups in a non-offensive way.

For a discussion on appropriative uses of STs, see among others Brontsema (2004), Croom (2014), or Bianchi (2014) who interestingly also provides a metalinguistic ("echoic") account of the defusing of their pejorative powers. I am tempted to conjecture that all non-pejorative instances of STs (embedded, appropriated and others) are ultimately metalinguistic.

The phenomenon of metalinguistic negation has been the focus of a lot of attention in the literature, but there exists to my best knowledge no consensual theory of the phenomenon[78].

My current purpose is not to develop a general theory, but rather to show that non-offensive uses of STs belong to this class of uses whose very existence is not debated. Recall that TCA takes non-projectability data to show that STs pejorativeness has a truth-conditional nature.

In (142)-(148) for instance, the availability of non-pejorative readings could be seen as evidence that different truth-conditional operators successfully affected the pejorative content of STs[79]:

(142) (!)There are no kikes, only Jews.

(143) (!)No jews are kikes.

(144) (!)There are no chinks at university, only chinese people.

(145) (!)Yao Ming is Chinese, but he's not a chink.

(146) (!)There are lots of Chinese people at Cal, but no chinks.

(147) (!)Chinese people are not chinks.

(148) (!)There are no chinks; racists are wrong.

I suspect, *contra* Hom and May, that the non-offensive readings of the above data are actually metalinguistic - and not propositional. Let us first consider the following informal characterization of metalinguistic negation. Horn (1985) describes metalinguistic negation as

> a device for objecting to a previous utterance on any grounds whatsoever - including its conventional or conversational *implicata*, its morphology, its style or register, or its phonetic realization (Horn 1985, p. 121)

[78] About the variety of metalinguistic uses, from clear cases of reported speech to less clear cases of mixed or perspectival uses of expressions, see for example Horn (1985,1989), or Recanati (2001, 2007).

[79] The examples are taken from Hom (2008) and Hom and May (2013).

that is, not as an device that would operate at the propositional level by reversing the truth-value of a proposition (or selecting the complement of a predicate), but as one that would operate at the level of discourse representations[80].

I intend to show that non-projectability does not confirm in itself that the pejorative content of STs is affected by *propositional* negation, only that it can be affected by *some use* of negation; so at least these data are not decisive for TCA. There is an alternative candidate.

But I make a stronger claim. I do not only claim that there is an alternative explanation to account for the non-pejorative readings of (142)-(148), I also argue that non-projective readings in (142)-(148) are in fact the result of a *metalinguistic* interpretation of negation, by considering linguistic constructions in which metalinguistic negation is *not* available.

I consider three such cases: i) the prefixal incorporation of negation, ii) the "it's false that" construction and iii) the "without being" construction[81].

6.3.1. "Non-F"

[80] It is no surprise to find in Horn (1985) an example of metalinguistic negation involving embedded STs: "It is relevant that metalinguistic negation can be employed by a speaker who wishes to reject the bigoted or chauvinistic point of view embodied in an earlier statement within the discourse context:

> (c) I beg your pardon: Lee isn't an 'uppity {nigger/broad/kike/wop/...}' - (s)he's a strong, vibrant {black/woman/Jew/Italian/...}. [...] 'I'm not 'colored' - I'm black!', 'I'm not a 'gentleman of the Israelite persuasion' - I'm a Jew!' " (Horn 1985, p. 133, footnote 10).

[81] Horn proposes another such test having to do with the distribution of polarity items (Horn 1989, pp. 370, 374, 396). According to Horn, metalinguistic negation shouldn't licence negative polarity items (e. g. "any"), nor inhibit positive polarity items (e. g. "some"). I prefer not to include such test because of the controversial and conflicting results that it produces (see e. g. Geurts 1998, p. 278).

Horn (Horn 1985, p. 140) observes that when we try to incorporate negation prefixally, metalinguistic readings are blocked. Observe what happens with "Yao Ming is not married": it is ambiguous between a propositional and a metalinguistic reading. In what follows, the material in parenthesis is only meant to help the reader access the relevant readings, they are not part of utterances.

(149) a. Yao Ming is not married. AMBIGUOUS (Prop. vs. Meta.)

b. Yao Ming is not married. (he is single) PROPOSITIONAL

c. Yao Ming is non-married. (he is single) PROPOSITIONAL

d. Yao Ming is not married. (he is joined in holy matrimony) METALINGUISTIC

e. ??Yao Ming is non-married. (he is joined in holy matrimony) ---

The test shows that when negation is incorporated, the only reading available is the propositional one (149c). Now consider the same test with a ST instead of the predicate "married":

(150) a. !Yao Ming is not a chink[82]. AMBIGUOUS (Prop. vs. Meta.)

b. !Yao Ming is not a chink. (he is Russian) PROPOSITIONAL

[82] Note also that the comparison of noun ST ("chink") with an adjective ("married") does not harm the point. The results replicate with adjectival STs like "queer" (but the well-known appropriative use of "queer" introduces unneeded complexity):

a. John is not queer. AMBIGUOUS

b. John is not queer. (he is heterosexual) PROPOSITIONAL

c. John is non-queer. (he is heterosexual) PROPOSITIONAL

d. John is not queer. (he is homosexual) METALINGUISTIC

e. ??John is non-queer. (he is homosexual) ---

c. !Yao Ming is a non-chink[83]. (he is Russian) PROPOSITIONAL

d. !Yao Ming is not a chink. (he is Chinese) METALINGUISTIC

e. !??Yao ming is a non-chink. (he is Chinese) ---

Once more, (150a) is ambiguous between a propositional and a metalinguistic reading of negation, i. e., it conveys (or not) a pejorative content towards Chinese people depending on the interpretation of negation (readings (150b) and (150d) respectively).

The prefixal incorporation of negation in (150c) encourages the internal reading of negation, under which the ST conveys a pejorative content towards Chinese people. On the other hand, if we encourage a metalinguistic interpretation of negation in the presence of incorporated negation, the result is deviant (150e)[84].

Overall, this is evidence that the non-pejorative reading of (150a) is the result of a metalinguistic effect, rather than truth-conditional computations.

6.3.2. "It's False That"

Constructions like "it's false that" also block metalinguistic readings of negation (Horn 1989, p. 416):

(151) a. It's false that John is married. (he is single)

[83] Note that English does not seem to have STs with already incorporated negation, but that it is harmless to use a neologism in this case. Imagine a situation in which speakers divide the world between people who are Chinese and people who aren't. They could here apply the predicates "Chinese" and "Non-Chinese". Racists would similarly apply the pejorative predicates "Chink" and "Non-Chink". It is harmless to use a neologism, as what matters in this case is the contrast between c-cases and e-cases.

[84] Väyrynen (2013) applies the same test to thick terms, which, according to B. Williams (1985)'s definition, "mix classification and attitude" (e. g. "reckless" or "brave").

b. ??It's false that John is married. (he is joined in holy matrimony)

From the ambiguous "John is not married", only the propositional reading survives when negation takes the above form, just like when it is incorporated. Now compare:

(152) a. !It's false that John is a kike. (he is catholic)

b. !??It's false that John is a kike. (he is Jewish)

Again, an utterance of "It's false that John is a kike" is interpretable only if the intended reading is propositional (152a); it is not felicitous with a metalinguistic reading.

Just like above, from the ambiguous "John is not a kike", the metalinguistic reading does not survive when negation takes the explicit propositional form. This is new evidence that non-pejorative readings of embedded STs (non-projectability data) are not the result of a truth-conditional interaction, and thus undermines TCA.

6.3.3. "Without Being"[85]

"Being F without being G" is just another way of "being F and not G". So when one says that "John is not married, he is F", we could as well say that "John is F without being married". This construction of negation also seems to rule out metalinguistic readings. Take the following pair:

(153) a. John lives with his partner, he is not married. PROPOSITIONAL

b. John is joined in holy matrimony, he is not married. METALINGUISTIC

And put it under the "without being" format:

(154) a. John lives with his partner, without being married

b. ??John is joined in holy matrimony, without being married.

[85] Benjamin Spector (p.c.) suggested this test to me, for which he credits Danny Fox.

Again, STs pattern alike:

(155) a. !John works in Israel, he is not a kike. PROPOSITIONAL

b. !John is Jewish, he is not a kike. METALINGUISTIC

(156) a. !John works in Israel, without being a kike.

b. !??John is Jewish, without being a kike.

Once more, an utterance of a ST under a negation of the form "without being" is interpretable only if the intended reading is propositional (156a, which is racist); it is not interpretable under a metalinguistic reading (156b). The metalinguistic reading of (155b) does not survive the "without being"-transformation. This is further evidence that non-pejorative readings of embedded STs (non-projectability data) are not the result of a truth-conditional interaction.

Taking stock, the three tests show that non-offensive uses of embedded STs pattern with metalinguistic uses of negation. Thus, non-projectability data like (142)-(148) do not show that the pejorative content of STs is propositional, as it is not cancelled by propositional negation but rather by *non-propositional*, metalinguistic negation. STs are not strictly speaking *used* in these cases[86].

[86] See Panzeri & Carrus (2016) for recent experimental work on this issue. The authors' findings suggest that negation is a special case of embedding as to the perception of pejorativeness, and that such a contrast might be due to the availability of a metalinguistic interpretation.

There is a potential worry about the above argument that it is useful to address. Here is the potential objection: failure to incorporate - and other such tests - shows only that negation is not *internal*; it does not show that negation is *metalinguistic* as there is a third option: *external* negation. I argue that this objection does not obtain.

In cases with singular terms, internal negation and external negation give rise to one and the same reading. The following example illustrates the distinction between internal and external negation (both of which are propositional, not metalinguistic):

(157) All citizens are not armed.

Just like above, there are at least two readings of (157), depending on how the hearer interprets negation. An *internal* interpretation of negation gives rise to a reading under which no citizen carries a weapon; whereas an *external* interpretation of negation brings about a reading under which at least one citizen does not carry a weapon[87].

None of these two readings has to do with metalinguistic negation, as they have different truth-conditions and are easily modeled in terms of propositional negation only:

(158) All citizens are not armed.

Internal reading: $\forall x(Sx \rightarrow \neg Bx)$ (All citizens are non-armed)

External reading: $\neg\forall x(Sx \rightarrow Bx)$ (it's not the case that (all citizens are armed))

The ambiguity here has not to do with the *nature* of negation (propositional vs. metalinguistic), but with its *scope*. When negation is interpreted locally as taking scope over

[87] Discussing Russell's (1905) famous example, Horn remarks that the "so-called 'external' or 'marked' negation is often exemplified by the reading of 'The King of France is not bald' which is forced by the continuation … 'because there is no King of France' and which is true if France is a republic; by contrast, the 'internal' reading is either false or lacks truth value." (Horn, 1985, p. 121)

the predicate B, we obtain an internal reading. When it is interpreted higher in the structure as taking wide scope over the whole proposition, speakers derive an external reading[88].

Both the internal and the external readings are compatible with a truth-conditional/propositional analysis of negation, one need not calling on metalinguistic negation or other discourse level phenomena to account for the two readings. Note indeed that if we try to incorporate negation to (157), the external reading is lost, just as in the previous examples:

(159) All citizens are unarmed.

Internal reading: $\forall x(Sx \rightarrow \neg Bx)$ (All citizens are non-armed)

*External reading: $\neg\forall x\ (Sx \rightarrow Bx)$ (it's not the case that all citizens are armed)

Consequently, we could have the following worry: the above tests which show that negation fails to incorporate under non-pejorative readings of STs do not establish in themselves that negation is metalinguistic, because failure to incorporate only rules out *internal* propositional negation. The objection would stress that there are two candidates left: metalinguistic negation and *external* propositional negation, with only the former being incompatible with TCA.

This objection does not succeed because in the cases at stake, internal and external negation give rise to one and the same reading, and can thus not explain the observed ambiguity. Consider again the ambiguity of utterances as (160a):

(160) a. !John is not a faggot. AMBIGUOUS

We know thanks to (160b) and (160c) that at least one of the two readings is internal, that is, the one according to which John is non-F.

(160) b. !John is a non-faggot. (he is not homosexual) PROJECTION

c. !John is not a faggot. (he is homosexual) NO PROJECTION

[88] It is as if propositional negation came in two varieties and could be respectively computed at different depth: the one selects a predicate and renders its absolute complement set, the other selects a proposition and transforms its truth-value.

An external interpretation of negation in this case corresponds to a reading under which it is not the case that John is not F. But for any F, there is no truth-conditional difference between John's being a non-F and its not being the case that John is a F.

In other terms, when singular terms are involved (as opposed to other determiners, like quantifiers as in (157)), the different structures interpretable depending on the scope of negation do not give rise to different truth-conditions. Therefore, the non-internal reading of (160a) cannot correspond to an external reading: it is metalinguistic[89].

[89] Note that Hom and May also put forward non-negative non-projectability data, such as:

a. !People treating Jews as kikes are anti-Semitic.

b. !Max doubts that Jews are kikes.

c. !Racists believe that Chinese people are chinks.

d. !Thinking that Chinese people are chinks is to be radically wrong about the world.

I will not discuss these cases in detail. I just remark that they all involve propositional attitude expressions, and are therefore good candidates for triggering other perspectival effects (on this topic, see Horn 1985, Carston 1996, Geurts 1998, Pitts 2011, Recanati 2001, 2007).

TCA makes another unwelcome prediction in the case of a special use of conditionals, where the patent falsity of the consequent implicates the falsity of the antecedent (call it the Absurd Counterfactual Conditional, or ACC[90]).

ACCs can be seen, roughly, as a way to suggest that the antecedent is as improbable as the consequent. (161) illustrates such a case, where the inference that the speaker believes in unicorns does not project:

(161) If Mary saw a unicorn, then I'm the queen of England.

NO PROJECTION (no inference that the speaker believes in unicorns)

An ordinary speaker (a speaker who doesn't believe in unicorns, and who is not the queen of England) would not typically utter (161) to express her belief of a counterfactual dependency between Mary's seeing a unicorn and herself being the queen of England. Instead, she would use the obvious falsity of the consequent to let hearers infer that she believes that Mary could *not* have seen a unicorn, that it is impossible (or at least that as improbable as the consequent).

According to TCA, it should be possible to use an ACC in order to convey one's dissociation from the fiction of racism. Nevertheless, slurs don't pattern in this way in ACCs. Compare the behavior of "homeopathy" and "nigger" in an ACC:

(162) If homeopathy cured Mary, then I'm the queen of England.

NO PROJECTION (no inference that the speaker believes in homeopathy)

(163) !If a nigger cured Mary, then I'm the queen of England.

PROJECTION (inference that the speaker is racist)

[90] The discussion of such data first came up in conversations with Benjamin Spector.

The obvious falsity of "I'm the queen of England" uttered by anyone but the actual queen of England suggests that the antecedent is false, that is, (162) does not trigger the inference that the speaker believes in homeopathy: it does not project.

And yet, the pejorative content of (163) projects. If TCA was right about STs, a speaker could utter (163) with the intention to convey her belief that there is not such a thing as a "nigger". But that is impossible: where (162) can be used to express one's disbelief towards homeopathy, (163) *cannot* be used non-pejoratively to express one's disdain towards racism.

In the next section, I consider and object to Hom and May's attempt to accommodate the intuition that embedded STs still carry a pejorative content.

As we saw above, according to TCA, STs under negation (and other operators) are not derogatory, in the sense that they don't *predicate* negative properties of a subject. Nevertheless, Hom and May recognize that something generates the intuition that the pejorative content of STs does project under semantic embedding. Consider again:

(137) !Obama is not a chink.

(138) !There are no kikes at my office.

(139) !If Freddie Mercury was a fag, then he was worthy of contempt.

Hom (2012) and Hom and May (2013) distinguish the phenomenon of "derogation" - predication of a negative moral property based on the subjects belonging to a group - from another phenomenon, somehow similar in its effects. In addition to being derogatory, the use of STs would generate what they call "offense". As Hom puts it,

> (…) [derogation] is an objective feature of the semantic contents of pejorative terms. Derogation is the result of the actual predication, or application, of a slur or pejorative term to its intended target group. (...) [Offense] is a subjective effect of the semantic contents of pejorative terms in a context. Offense is a psychological result on the part of the discourse participants, and is a function of their beliefs and values. (Hom 2012, p. 397)

Under this first characterization of offense, there are non-linguistic factors at work, such as our values and beliefs, in our perception that such or such utterance is pejorative. According to Hom (2012) and Hom and May (2013), this intuition stems from a confusion between "offense" and "derogation".

There would thus be two possible factors responsible for the negative effects that STs elicit[91]: "derogation" and "offense". With these two parameters, an utterance "U" featuring a

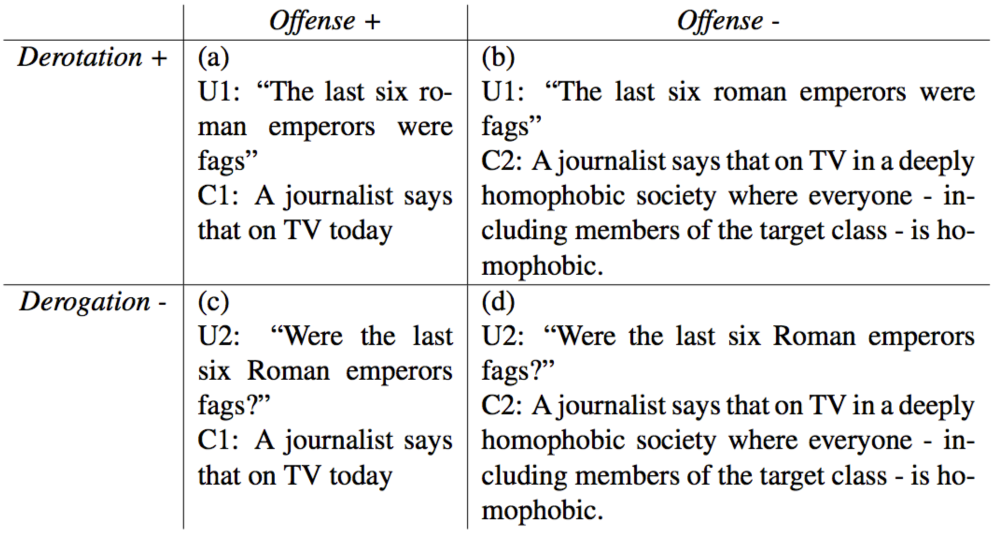
[91] Jeshion (2013b) notices that the distinction makes the surprising prediction that "John is a nigger" and "Is John a nigger?" are disparaging for completely different reasons, which at least calls for clarification. Furthermore, it predicts that "Max is not a chink, he is a nigger"

ST "S" in a context "C" can stand in four possible states depending on the linguistic environment and the conversational context, as illustrated in Figure 1.

Whenever negative moral properties are predicated of a subject (like in U1), there is *derogation* (in both cases a and b); on the other hand, *offense* is brought about only in contexts of utterance where a "psychological result" is expected, in particular, when it is problematic in the context of utterance to suggest that the members of the target class are despicable, or anything along those lines (cases a and c); otherwise - for example in a deeply homophobic society where everyone - including members of the target class - is homophobic, offense does not arise (cases b and d).

So for example according to Hom and May a question like U2 *does not* carry any disparaging content towards homosexual people: whether or not it is offensive depends on the context, and in the present hyperbolic-homophobic scenario, the question U2 is neither derogatory nor offensive (case d):

FIGURE 1. HOM AND MAY'S DEROGATION/OFFENSE DISTINCTION



	Offense +	*Offense -*
Derotation +	(a) U1: "The last six roman emperors were fags" C1: A journalist says that on TV today	(b) U1: "The last six roman emperors were fags" C2: A journalist says that on TV in a deeply homophobic society where everyone - including members of the target class - is homophobic.
Derogation -	(c) U2: "Were the last six Roman emperors fags?" C1: A journalist says that on TV today	(d) U2: "Were the last six Roman emperors fags?" C2: A journalist says that on TV in a deeply homophobic society where everyone - including members of the target class - is homophobic.

is derogatory (in the technical sense) towards African-Americans, but not toward Chinese people. In particular, it predicts that the reasons why we feel that this utterance is disparaging these two groups are of a completely different nature.

Again, with this move, Hom and May's version of TCA has two different mechanisms able to account for the negative effects of STs: a truth-conditional component on the one-hand, and a psychological and pragmatically driven component on the other hand.

We have also observed that in this framework, there is a difference in pejorativeness between (b) and (d): for Hom and May, (d) is supposed to be non-pejorative, as there is no "derogation" nor "offense", whereas (b) is be pejorative. I find this prediction unsatisfactory.

However, Hom and May provide another, finer grained, analysis of the *causes* of offense, which is able to explain case (d):

> Offensiveness can be linguistically triggered, because when speakers use predicates, they typically conversationally implicate their commitment to the non-null extensionality of the predicate. (Hom and May 2013, p. 310)

The idea seems to work like this: speakers tend to use terms that they believe have a non-empty extension; this is true of any predicate (tables, bottles, etc). For example, if John asks Mary whether she ever speaks to angels, Mary and bystanders will conversationally implicate that John believes that angels exist[92].

Likewise, if John asks Mary whether she ever speaks to wops, Mary and bystanders will tend to infer that John believes that "wops" exist (i. e. that there are people who are bad because of being Italian). Hence our intuition of projection: any use of STs, embedded or not, trigger an implicature of non-null extensionality, and that is offensive, given the alleged meaning of STs.

Nonetheless, relying on non-vacuity inferences to explain the pejorativeness of embedded STs is inadequate. Take a construction where non-vacuity inferences are usually blocked, like "there is no F". We do not infer that the speaker believes in the existence of God from her utterance of "there is no God".

[92] Conversational implicatures are expected to be cancellable. Hom and May call here on non-projectability data to argue that indeed, *offense* (triggered by a conversational implicature) can be cancelled, like in "There are no kikes, kikes don't exist", "John is not a kike because there is no such a thing" or "No Jews are kikes". I have just argued that these cases are better understood in terms of metalinguistic effects.

The same holds for "there are no vampires". Now, note that although vampires don't exist, there are STs for vampires (e. g. "fangs"), just like for other fictional entities ("pointyear" for elves, "toaster" for robots, "furface" or "moondog" for werewolves[93]). Imagine that Mary wants to reassure John, who is afraid of being bitten by a vampire on his way home. She could utter (164) or (39):

(164) Don't worry! There are no vampires! They don't exist.

(165) !Don't worry! There are no fangs! They don't exist.

Although neither (164) nor (165) trigger existential inferences, the utterance of (165) still carries Mary's negative evaluation of vampires. The negative evaluative content about vampires in (165) cannot be the result of non-emptiness inferences[94]. Therefore, non-null-extensionality implicatures do not explain the projection of the evaluative content of STs[95].

[93] The "fangs" example comes from the HBO tv-series True Blood; among gamers, there are slurs for all sort fictional entities (tvtropes.org/pmwiki/pmwiki.php/Main/FantasticSlurs).

[94] Note that in the previous section I have already considered a construction that *can* be used in a way that non-null-extensionality implicatures are blocked, namely ACCs.

[95] We could instead propose that what triggers offense is not the inference of non-emptiness *per se*, but the inference of *possible* non-emptiness. "Wop" triggers offense because it suggests that the speaker believes it is *possible* that Italians ought to be the target of negative moral evaluation because of being Italian. This variation won't work either, as Mary could as well try to reassure John in saying "Don't worry, there could be no fangs! It is simply impossible that fangs exist".

I take it that what a term (or concept) *t* in a language (or language of thought) *L* means or does not mean is an empirical matter, i. e., I take a naturalistic stance on semantics. The methods of semantics consist in investigating competent speaker's judgments (usually of truth and falsity, or of appropriateness) about various utterances in various contexts, in order to probe the pre-existing meaning components of linguistic expressions and constructions (which are not necessarily transparent).

Now, the hypothesis that BOCHE means something along the lines of "ought to be the target of negative moral evaluation because of being German" should be supported by evidence. I have claimed that the readings under which utterances like "there are no kikes" are non-disparaging do *not* count as positive evidence for TCA, as they are best analyzed as metalinguistic interpretations of the involved concepts.

Then, what could further motivate the hypothesis that the pejorative content of slurring concepts is truth-conditional? I speculate that the underlying motivation for such analysis is to formulate a theory such that the following normative requirement is met: utterances and thoughts such as "John is a chink" must be false, or at least not truth-apt. If this normative condition is really what motivates TCA, I can make two observations.

First, we shall distinguish between a technical and a folk notion of "truth". Under the technical use of the truth predicate, as it is used in formal semantics (at least since Tarski), for an utterance or thought to be true just means that the world satisfies certain conditions, conditions whose nature depends on the conventional encoded properties of the linguistic/mental items involved and the way they are put together. Under this understanding of the truth predicate, an utterance being *true* does not entail that it is *acceptable* (e. g. adults speakers tend not to accept "some elephants are mammals" even if it is literally true (Bott and Noveck 2004[96])).

[96] Sennet and Copp (2015, p. 1091) make a similar point in emphasizing the distinction between *truth* and *felicity*.

It should therefore not frighten theorists to discover truth-conditions such that utterances and thoughts of "John is a boche" come out *true*. For example, if BOCHE and GERMAN are coreferential, "John is a boche" is true only if John is German, and the pejorative content triggered by uses of STs/deployments of SCs might still be managed by other dimensions of the linguistic machinery (implicatures, presuppositions, etc).

The second clarification has to do with the role and methodologies of semantics as an empirical discipline. We shall first carefully recall that TCA (just like TCSE and TCE) can be separated in two distinct claims. It contains a *semantic* claim - according to which a pejorative S targeting individuals G means something along the lines of "ought to be the target of negative moral evaluation because of being G" - and a *moral* claim - according to which no one ought to be the target of negative moral evaluation because of being G.

The empty-extensionality thesis follows from the moral claim combined with the semantic claim, that is, given the content of KIKE, and given the moral facts, the concept KIKE is not instantiated: it renders an empty extension for any point of evaluation (that is, the intension of "kike" is a function mapping worlds to the empty set, just like "square circle[97]").

If BOCHE really meant "bad because of being German", an utterance/thought of "John is a boche" would indeed be false at all worlds in all contexts, because the moral fact obviously holds. But the initial question was not "what would be the truth-value of 'John is a boche' if BOCHE meant X?"; the real question is "what does BOCHE mean[98]?". And the task of semantics is to investigate the truth-conditions and other properties of real utterances and thoughts containing "boche" or BOCHE, and other such terms and concepts.

The debate about slurring concepts might require an investigation on the relation between the research on meaning and normative desiderata (such as the desideratum that "John is a kike" *cannot* be true).

[97] And not like "golden mountain": in some worlds, the golden mountain might exist, but there are no accessible world in which a square circle exists.

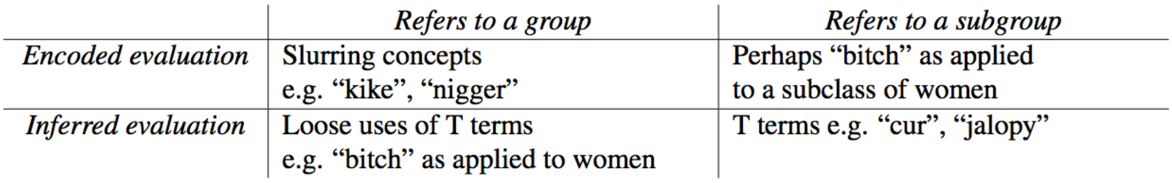
[98] Sennet and Copp (2015, pp. 1090-1092) raise similar considerations in discussing Hom and May's "conceivability" argument and "Frege cases".

In the following chapters, I am going to sketch other non-hybrid accounts, which do not face this problem and does not have the consequence that STs have an empty extension. The account I am going to investigate, based on the notion of response-dependence, maintains that SCs have NCs.

Another advantage of the new account I am going to propose and evaluate is that it gives pride of place to the emotional dimension which all the accounts I have considered so far (with the exception of the speech act account) fail to properly take into account.

Summing up before moving on, the discussion in this and the previous chapter focused first on whether S-terms and T-terms refer to the same, or to a narrower, class than their counterpart, as well as on the conventional/non-conventional character of their evaluative content. We ended up with different dimensions involved in STs and T-terms, and can eventually characterize SCs as follows:

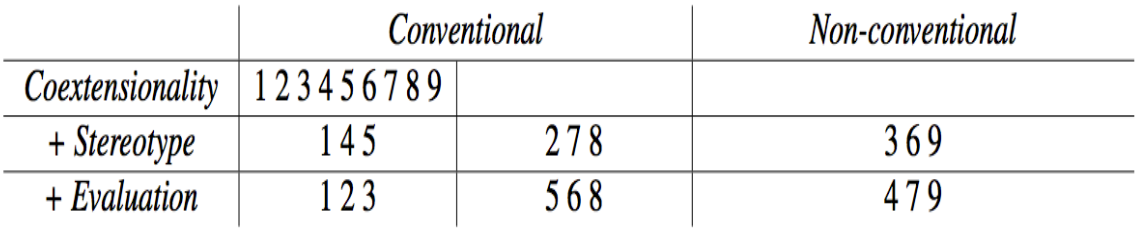
FIGURE 2. FOUR CLASSES OF EVALUATIVE TERMS AND CONCEPTS



	Refers to a group	*Refers to a subgroup*
Encoded evaluation	Slurring concepts e.g. “kike”, “nigger”	Perhaps “bitch” as applied to a subclass of women
Inferred evaluation	Loose uses of T terms e.g. “bitch” as applied to women	T terms e.g. “cur”, “jalopy”

With two dimensions of meaning (conventional and non-conventional) two ways of being conventionally encoded (truth-conditionally or in an expressive, CI dimension), and three aspects relevant to STs (their relation with a counterpart, their evoking a stereotype and their evaluative import), we obtain the following table, representing nine conceivable positions whose main variations we discussed here and earlier:

FIGURE 3. NINE VIEWS ON SLURRING TERMS AND CONCEPTS



	Conventional		*Non-conventional*
Coextensionality	1 2 3 4 5 6 7 8 9		
+ Stereotype	1 4 5	2 7 8	3 6 9
+ Evaluation	1 2 3	5 6 8	4 7 9

1: à la Hom (2008)

2-3: Variations on Hom and May (2013)

4: Loose uses of reference-based thick terms

5: Reference-based view of thick concepts after conventionalization

6-7-8: Variations on hybrid expressivism

9: à la Nunberg (2016)

So we see in figure 3, several variations on hybrid expressivism are put forward in the literature. For instance, Camp (2013) proposes, among other hybrid expressivists, that

> slurs conventionally signal a speaker's allegiance to a derogating perspective on the group identified by the slur's extension-determining core. (Camp, 2013, p. 331)

As we will see now in the later chapters, it is possible to give an account of slurring concepts in a more detailed and precise manner than just saying that "their evaluative content is conventionally associated", or that "they signal a derogating perspective".

What is the nature of this "conventional association"; or of this "derogating perspective"? The following chapter, singling out response-dependent concepts as a potentially relevant class of concepts to identify SCs with, can be seen as an attempt to characterize the notion of a "derogating perspective".

Chapter 7. A Response-Dependent Account

We saw that Frege's distinction between *sense* and *tone*, Grice's distinction between *what is said* and what is *conventionally implicated*, along with the modern distinction between the *descriptive* and the *expressive* dimensions - we saw that all three seem to link a reference-fixing component and an expressive component somewhat arbitrarily, as if the two were independent from each other and conjoined by convention.

Still facing the need to introduce a link between the descriptive and the expressive dimensions of slurring representations, we just examined two views attempting at doing so. The first one, based on the literature about "thick terms", was not fully satisfactory (because of a possible ambiguity of the terms, and of a distinctive pattern regarding redundancy judgments. The second - more radical - located all main dimensions of slurring concepts (SCs) in their truth-conditional dimension, but not fully satisfactory mainly because of its wrong predictions regarding projection.

I will now turn to an investigation of another family of theories introducing a strong link between the descriptive and the expressive dimensions: response-dependent accounts of slurring concepts. The idea underlying the assimilation of slurring concepts to response-dependent concepts such as RED and other secondary quality concepts is that all these concepts seem inherently tied to non-conceptual states. Slurring concepts would be grounded on a certain cognitive non-conceptual response to certain (clusters of) properties, and in that sense be similar to concepts such as RED.

That a concept is a response-dependent concept has two important consequences: it imposes possession conditions on the concept, and it provides a theory of reference determination for the concept. Indeed, a subject can be said to possess a response-dependent concept when she is suitably related to a particular non-conceptual state. There are two main types of relations that are suitable. First, she might directly experience the non-conceptual state herself. Second, she might indirectly know the role that the non-conceptual states play in others. The concept is somehow dependent on that non-conceptual state, just like color concepts might be dependent on color perceptions.

Second, the concept secures its reference through the non-conceptual cognitive state it is built on, picking out the objective properties that are responsible for its triggering, just like a particular color concept refers to a particular set of physical properties that cause a particular color perception.

To develop the view, I will proceed as follows. First, I will introduce the general debates surrounding response-dependence, the metaphysical and epistemological issues that it is usually meant to address, and the realist picture of the world that usually accompanies it. I will then discuss in more detail the notion of response dependence, introducing response-dependent biconditionals and making a few necessary clarifications about their functioning and range of application.

That in place, we will then be able to apply the notion of response-dependence to slurring concepts, that is, to investigate response-dependent accounts of slurring concepts (RDA). I will do so in two steps. First, I will develop an account I coin "RDAred", ensuing from an analogy with secondary quality concepts such as RED.

I will introduce two meta-semantic distinctions: a distinction between opaque and transparent cases of response-dependence on the one hand, and between reflexive and non-reflexive cases on the other hand. The first distinguishes cases as a function of the subject's access to the response-dependent biconditional governing the concept. Indeed, subjects are not necessarily aware that their concept involves their own response.

The second distinguishes cases where concept possession is possible without the response from cases where that is not possible. It follows from RDAred that only responders normally possess slurring concepts, and that non-responders are likely to have a deferential concept. We will see that with the properties of opacity and reflexivity, RDAred treats SCs as a kind of indexical concepts.

7.1. From Error Theories to Realism

In order to introduce the notion of Response-Dependence that I will use to attempt to model slurring concepts, let us take a closer look at a paradigmatic type of response-dependent predicate that raises serious epistemological and semantic issues.

Among response-dependent predicates, color predicates (e. g. "red", "blue") constitute a paradigmatic case, but predicates based on other modalities (e. g. "loud", "spicy", "heavy" etc. all expressing what are often called "secondary qualities") raise similar questions, as do moral predicates (e. g. "right", "impermissible") evaluative predicates (e. g. "good", "disgusting"), thick evaluative predicates (e. g. "chaste", "lust" etc. see chapter 4), some gradable or vague predicates (e. g. "expensive", "tall") etc.

There are several positions one could be willing to take with regard to these predicates, and I will now introduce the general landscape of these different positions before getting back to slurs. I make this detour because the semantic analysis of a large range of derogatory expressions and other expressives might face the same kind of conundrums as color and other secondary quality concepts. I will eventually argue that an assimilation of the concepts expressed by slurs (SCs) to a wider class of subjective or partly-subjective concepts (more precisely to response-dependent concepts) constitutes a first plausible general explanation of pejorative thought and talk.

In particular, this detour through questions about the nature of such concepts will eventually lead us to explore the view that slurring expressions have an indexical component, that is, to a first approximation, that they refer to objects and individuals that are in a particular relation with the speaker herself. I will then discuss a series of issues and limitations of that view, so as to move towards what I think is a more adequate account, based on the notion of (psychological) essentialism (chapter 10).

7.1.1. Color Science and Error theories

What happens when we think or say that ripe tomatoes are "red"? Do we think or speak truly? What do color terms and color concepts refer to, if anything? What is the content, or cognitive value, of color concepts? Color science seems to tell us that there are simply no physical properties that satisfy the requirements for being colors, corresponding to the way that normal humans use color terms to ascribe color properties to objects. Colors don't exist in objects, we are told; they are just present in our experience. Any view denying existence to color can be called an error theory of color[99].

Different scientific methods can be used to investigate the nature of color. Chemistry and physics can study the character and composition of light and physical properties of surfaces, cognitive neuroscience can study the physiological and cellular mechanisms underlying color perception, and cognitive psychology can study the mental organization of the colors in a systematic fashion with different dimensions (e. g. hue, value or chroma).

Science informs us that our cognitive system somewhat artificially divides a continuum of physical properties into color categories, artificially opposing blue and yellow on the one hand and green and red on the other hand. It also informs us that color cannot be identified with reflectance (see e. g. Hurvich & Jameson 1957).

Also, research on animal color perception suggests that color is relative to species: for instance bees see ultraviolet radiations but cannot distinguish red from black (Von Frisch 1950). So the advances in color science suggest that color cannot be a completely objective categorical property. It is inherently tied to perception.

In fact, since the seventeenth century, many (color) scientists and philosophers have been led to hold different varieties of error theories, claiming that color predicates don't refer (directly) to an objective reality, but rather that humans "project" on reality some features of the cognitive system itself. For instance, Hume (1739) wrote that

[99] Error theories are thus antirealist views about colors.

> sounds, colors, heat and cold, according to modern philosophy are not qualities in objects, but perceptions in the mind. (book III, part I, sect. 1, p. 177).

Seeing an object as red, under this view, consists in projecting onto the world a color produced by the perceiving system itself, just like an illusion. That is also the conclusion reached in the following quotation from the classical *Vision Science* by Palmer (1999):

> People universally believe that objects look colored because they are colored, just as we experience them. The sky looks blue because it is blue, grass looks green because it is green, and blood looks red because it is red. As surprising as it may seem, these beliefs are fundamentally mistaken. Neither objects nor lights are actually "colored" in anything like the way we experience them. Rather, color is a psychological property of our visual experiences when we look at objects and lights, not a physical property of those objects or lights. (p. 95)

Considering these observations[100], it seems as Dennett (1993) puts it that "modern Science [has] removed the color from the physical world, replacing it with colorless electromagnetic radiation of various wave-lengths" (p. 370) encountering surfaces that reflect and absorb that radiation which then hit the eye of the observer whose visual system is responding in certain manners so as to give rise to color experience and then to color concepts[101].

[100] There are numerous similar remarks in the literature:

"But if we wish, so to speak, to detach them from us and invest the objects with them, then we have no idea what we are doing. We find ourselves attributing them to objects only because, on the one hand, we must suppose they are caused by something, and because, on the other, their cause is altogether hidden from us" (de Condillac 2001, p. 10)

"...the intentional content of visual experience represents external objects as possessing colour qualities that belong, in fact, only to regions of the visual field. By 'gilding or staining all natural objects with the colours borrowed from internal sentiment', as Hume puts it, the mind 'raises in a manner a new creation' ". (Boghossian and Velleman 1989, p. 96)

[101] Note that for my present purpose, I don't need to take a stand on what color experiences consist in (on that matter see for instance Chalmers 1995, or O'Regan and Noë, 2001).

7.1.2. Alternatives to Error Theories

A direct and unwelcome consequence of error theories is that we are constantly making mistakes when we ascribe color properties or moral properties to objects and events in the world around us (Joyce, 2009). Color terms and concepts would be empty, in the sense that their reference would be a property that nothing instantiates (taking on board Frege's distinction between extension and reference). This is unwelcome, because if one is trying to understand the semantics of color concepts, there is not much more we can do than start from the subjects' applications of color concepts and their intuitions on truth and falsity.

Imagine that extraterrestrial intelligent creatures were to land on earth, and tried to figure out the functioning of the noises we make. They might end up wondering what "chair" means. I fail to see what more they could do than look at what objects or properties speakers ascribe the word "chair" to, and this is just another way of investigating speaker's intuitions on truth and falsity.

"This is a chair" will be judged true by most speakers when presented with a chair, and false when presented with, say, an apple. The creatures will also figure out that "chair" is vague, that there is some contextual factors and so on. Now, when one is trying to give a semantic analysis of color terms, as of any other terms, one is arguably just in the same position as the extraterrestrial creatures of our toy example (for more on this issue, see the radical translation scenarios in Quine 1960).

The first step is to look at how people use the term, what objects or properties they (intend to) apply it to, and so on. The theorist might not know whether it is true or not that ripe tomatoes are really red, whether it actually makes sense to ask such a question, whether the term "red" is in fact empty or not, or even whether it is relevant for the project of giving a semantics for color terms. But she does know that, in order to account for the meaning of "red", she can only start with speaker's judgments that "tomatoes are red" is true, and that "bananas are red" is false.

This does not exclude the possibility that some terms have an empty extension, but that may be discovered only after the theorists figured out the *intended* target of the terms and concepts she is studying. Only once the meaning of the term is known can we check whether the objects have the ascribed property or not. In order to establish that a term like "red" does not refer, one must first know what it is supposed to pick out.

For example, it is because we know that the noun "unicorn" applies to horses with a horn, and because we know that horses with a horn do not exist, that we know that "unicorn" fails to refer. Error theories of color seem to presuppose that color terms express a certain property that nothing in the world instantiates. But the goal is not (only) to say which properties are instantiated and which aren't, it is also to understand what properties color term denote, whether they are instantiated or not.

I claim that the conclusions drawn by Palmer and other color scientists do not in fact follow from what was discovered, because establishing that color concepts have empty extensions requires a semantic analysis of color concepts, and color science was not interested in semantic analysis in the first place. When one knows everything about the structure and functioning of the perceptual system, about the properties of light and surfaces and so on, one still does not know what the natural language term "red" means (or what it refers to, if at all).

Not only does emptiness not follow, it also has some unwanted consequences, as we just began to see. First, error theories entail that everyday color talk is truth-value-less, that one is constantly and systematically making mistakes in our everyday color judgments. But there is no need to go that far in order to be able to give an adequate account of color perception, thought and talk, while remaining consistent with scientific findings, as we shall see later.

Second, if color terms denote a property that is such that they have no extension, then colorblind people have a representation of their environment just as accurate (if not more accurate), as subjects with well-functioning perceptual systems. Since for error theories, color is an illusion, it is an illusion which colorblind people are not the victims of.

This, at least, calls for a serious reconsideration of simple error theories. Don't color-sensitive creatures get more information from their environment than color-blind ones? Isn't that information useful and adequate in many ways? And if it is, what is it about the term "red" or concept RED that makes it empty? Finally, if color perception is illusory, how can one characterize the difference between veridical and illusory perception of color?

Consider the example of the famous lilac chaser illusion, where one is presented with twelve magenta dots arranged in a circle around a fixation point. A pattern of flickering of the dots gives rise to the perception of a moving green dot (green is magenta's complementary color). In fact, although there are magenta dots on the screen, there are no green dots at all. The perception of magenta dots is in some sense less deceiving, more veridical or adequate, than the perception of the green dot(s).

Error theories of color seem to be under-equipped to account for this contrast, between what seems like veridical perception of (magenta) colors, and illusory ones (here of green). For error theories, the statement "There is a magenta dot" is just as false as "There is a green dot", and this violates the subject's' intuitions that one was trying to give a theory of.

Fortunately, one need not claim with Palmer that it is false that "the sky looks blue because it is blue, grass looks green because it is green, and blood looks red because it is red". What then is the best way to escape this unnecessary postulate of permanent and generalized errors?

We evoked earlier that noncognitivism claims that moral (or aesthetic) judgments like "Stealing is wrong" just have the appearances of a description of the world, but in fact do not have real propositional content (Blackburn, 1984, 199, van Roojen, 2014). Therefore, truth and falsity would not be the relevant dimension for assessing moral judgments (see Richard, 2008 for an analogous claim about slurs).

At least two versions of noncognitivism can be distinguished. Prescriptivists equate "Stealing is wrong" with the order "Don't steal!" (Carnap, 1935). On the other hand, expressivists state that "Stealing is wrong" expresses no more proposition than "Stealing, boo!" (Stevenson, 1935, Blackburn, 1993).

Another possibility is that these statements, even though they look like a description of an external reality, in fact express another proposition, about the utterer herself. I call this view the misplaced proposition view, as we discussed earlier. Under this theory, it is not that moral and aesthetic statements are not propositional, it is that the proposition which is expressed by them, for some reason, is not the one it seems to be on the surface. For instance, "That is beautiful" is equated with "I like that", or "That is wrong" with "I disapprove of that".

But noncognitivism about color would be almost unintelligible: what would the non-propositional, underlying content of color statements like "That is red"? Under a prescriptivist version of noncognitivism, one does not see what could be a suitable prescription to paraphrase the statement.

Indeed, given that moral predicates usually apply to actions of agents, a view equating moral statements with orders intended to influence the actions of agents can make sense. But color predicates do not apply to actions of agents, so it would be really counter-intuitive and tricky to try the prescriptivist line.

Similarly, the expressivist version would be odd. Again, there are expressives like "boo" or "hurray" which, somehow, target a dimension of some agent's behavior, and color really has nothing to do with agents. Maybe could one try to equate "Tomatoes are red" with "Tomatoes, red!", with red somehow expressing a perceptual state rather that describing an objective state of affair? It is hard to say, but it seems that the expressivist version of non-cognitivism is tailor-made for emotional (or moral) predicates, so that forcing perceptual ones into the picture looks somewhat artificial[102].

Finally, the misplaced proposition view of color would equate "that is red" with something along the lines of "I see that reddishly" (where the adverb corresponds to a property of the experience). This view would be consistent with what color science seems to have taught us: that we are wrong in our everyday color talk and thought insofar as we ascribe the property to the object, while it is a feature of our experience.

So the misplaced proposition view does not really allow us to escape the problematic conclusion that we keep making mistakes when we ascribe colors to objects. It would not be the objects that are red but the experience itself, contrary to the naïve view.

A better way of escaping that conclusion consists in treating the relevant properties not as qualities of experience, but as powers of the things in the world to provoke certain experiences in the minds of normal subjects. To be clear, although error theories claim that

[102] An answer could be that any kind of state could be expressed in the expressivist's sense, be it emotional or perceptual.

colors do not exist, they cannot claim that there is really nothing out there in the world accounting (at least partially) for color experiences.

When I open my eyes in front of a new scene, I suddenly see green over here, yellow over there, red here etc. That cannot be totally random. If I close my eyes and open them again, I will perceive the same colors. Also, others tend to use the same color terms as me to describe the same parts of the scene.

So even if the world really is colorless, so that color predicates fail to refer, still, there must be some properties of the scene that play a role in color perception. In that sense at least, color error theories are not purely error-theoretic. Color perception is not pure hallucination for example, where there really is nothing out there that is perceived.

Recognizing that color perception, hence color thought and talk, is not totally arbitrary is the first step towards a more adequate account of color experience, thought and talk. RED could refer to the power, or the dispositional property, of certain surfaces of physical objects to produce in observers an experience of red. The power or disposition is in the external object, but the experience, the response, is in the perceiver. We see in this notion of a disposition a way out of puzzles about these concepts we don't easily manage to fit into referentialist accounts of the world, such as color concepts.

On the dispositionalist approach, facts about colors and other secondary qualities of objects are inherently linked to a certain class of observers (in contrast to so-called primary qualities like size, shape, motion, number (Locke, 1690)), but importantly, this does not make them fictional. As we shall see, one can draw an essential, constitutive link between certain properties and observers without thereby being an anti-realist about these properties.

The link to observers can be conceived of in different manners. Dennett discusses the following analogy, which is telling in the present context (Dennett, 1993). In virtue of the meaning of the word "suspect", it is logically impossible that someone is a suspect unless someone actually suspected her of something. Mary can be worthy of suspicion, but it is only when someone actually suspects her that she becomes a suspect. If colors were the "suspect" sort of concepts, it would entail that the tomato in my fridge is not red until I open the door and see it.

Contrary to "suspect", someone can be "lovely" without having been observed by any person who would find her lovely. Still, the application of the predicate is linked to a class of observers. What makes a snake "not lovely" and John "lovely" has to do with people's sensitivity to snakes and to John, respectively. Importantly, these people need not be the speakers themselves. John can be "lovely" even though he might not be lovely to us, but might be lovely to another, maybe more entitled, observer. (166) for example is not a contradictory statement:

(166) I know that John is lovely, but I don't find him lovely.

If a speaker can unproblematically utter (166), it must be that she recognizes her failure to perceive a property that really exists, although that very property somehow depends on some observers. Even though "lovely", unlike "suspect", requires no actual observation, it seems to be intrinsically tied to a certain class of observers. It is this sort of link between a property and a class of observers that is relevant to secondary qualities: these concepts correspond to dispositions, and objects can have dispositions even if the dispositions in question are not actualized.

I will now see in more detail how the dispositional view can help us to defend a form of realism about secondary qualities.

7.1.3. The Dispositional Approach

Locke proposed that "sensible qualities" were powers to trigger "ideas" in us:

> The power that is in any body, by reason of its insensible primary qualities, to operate after a peculiar manner on any of our senses, and thereby produce in us the different ideas of several colours, sounds, smells, tastes, etc. These are usually called sensible qualities. (Locke 1690, II, VIII, 23)

There are two ways to conceive of a disposition, and two types of dispositional concepts corresponding to these two ways. The reference of the concept could be a higher order property, or a lower-order property physically realizing the higher order property.

Let us make that distinction clearer with an example. Take the dispositional concept FRAGILE as applying to objects with a disposition to break under a relatively low level of pressure, to keep things simple. Now, there are numerous physical, chemical etc. reasons why an object can be broken under pressure (e. g. thin and sensitive junctions, weak chemical bonds, sensitivity to temperature and so on). When a speaker ascribes the predicate "fragile" to, say, a glass bottle, her ascription will be veridical, under our hypothesis, if and only if a relatively low level of pressure would break it, and so independent of the real physical property deriving that outcome.

Whether the glass bottle would break because it is made of very thin glass, of because it was broken before and re-assembled with cheap glue, is irrelevant to the evaluation of the claim "that bottle is fragile". The lower order property (of e. g. having weak bonds) explaining the higher order property (here to break under pressure), I will call the "realizer"[103].

In the case of FRAGILE, the disposition is a higher order property that can be instantiated by several realizers, but that is not necessarily the case for all dispositional concepts. There is a logical possibility of having a concept like FRAGILE which would apply to a bottle disposed to break under pressure if it has the realizer property P1, but not if it has the realizer property P2, even though P2 equally makes the bottle breakable under pressure.

Whether there are or not such concepts for multiply realizable dispositional properties in natural language and thought is an open question. Color physicalism is precisely the attempt to identify colors (the reference of color concepts) with the lower-order realizers of the higher-order disposition to cause color experiences.

Color physicalism claims that color consists in complex, microstructural properties of surfaces and lights[104] (Byrne and Hilbert, 1997, 2003). This view was defended by several

[103] Also known as the "causal basis" for a disposition.

[104] "Primitivism" is another simple objectivist view about secondary qualities, which can also apply to moral judgments (Maund, 2012). Color primitivism defends the view that colors are simple, *sui generis*, unanalyzable properties of physical bodies (Watkins, 2005). This move drastically answers the question of the extension of such concepts: when talking about color, we are talking about objective, mind-independent, intrinsic but irreducible properties

philosophers (e. g. Reid 1822, Armstrong 1969, Matthen 1988, Hilbert 1987 to refer to just a few), and has faced major criticisms.

One of the main criticisms is that despite years of scientific research in that direction, we haven't found any physical property fully accounting for the colors we see in objects, in the necessary and sufficient way which color physicalism requires. Reflectance surely has an important role in color perception, but since the underlying causes of reflectance are complex and diverse (light, volume, scattering, surface, diffraction etc.), a unique surface color can be associated with multiple reflectance curves. A particular color doesn't imply a particular reflectance, that is, reflectance is not sufficient.

Moreover, famous illusions in color perception have shown a great influence of context (e. g. the checker shadow illusion and its many variants), or of background in our perception of color, so that a unique reflectance profile can be associated with multiple experiences of colors. A particular reflectance doesn't imply a particular color, that is, reflectance is not necessary. The physical conditions of color experience in fact appear to be of extremely varied and complex nature: one has not found any such necessary or sufficient conditions yet, and most think we won't find any. Hence the penchant for color anti-realism.

Now, first of all, the mere fact that no such property was found doesn't establish that there are none. But more importantly, there is a coherent way to hold to color physicalism by

possessed by many material objects. There is thus no illusion or error in color perception (Gert, 2006).

This view, which is a rather mysterious way to resolve the mystery, has been largely abandoned, mostly because it seems contradictory with what we do know thanks to color science: there *are* complex and micro-structural properties of surfaces that really do play (at least partially) a causal role in our experiences of colors.

It is easy to see how one could hold a similar view about "good" or "beautiful" or "expensive", and simply say that these are simply objective, mind-independent, *sui generis*, irreducible properties of objects. But such claims, although they give an easy theoretical way out of the puzzle, won't resist any experimental finding linking real properties with the deployment of these concepts.

under-specifying the necessary and sufficient realizer properties. Since we have systematic subjective responses, we can relativize the physical properties, and just say that they are the properties actually provoking these responses, however diverse and complex they happen to be (Johnston, 1992, 2004). This corresponds to the idea that color concepts are response-dependent, referring to physical properties. The trick is in characterizing the relevant physical properties in terms of the response they determine.

After all, why should we expect the human perceptual system to be sensitive only to properties as simple and easily captured by physical theories as a reflectance curve? Isn't sensitivity to extremely complex and diverse environmental features exactly what we should expect from a biological system anyway?

7.1.4. The Metaphysics of Response-Dependence

Response-dependent concepts target properties that objects have or don't have – the property of causing certain experiences, in certain subjects, under certain circumstances. This makes it possible to explain why judgments involving such concepts can be true or false. This is a great advantage over error theories. Error theories cannot explain how it can be true that a given object (say, a ripe tomato) is red. But it is a fact, if anything is, that a (normal) ripe tomato is red.

The response-dependent view captures that insight, so we don't have to buy color antirealism. We can resist error theories, with their devastating consequences: and to achieve that it is sufficient to concede that color is not a mind-independent property. A property does not have to be mind-independent to count as an objective feature of the world.

To show that such a position is not metaphysically problematic, I suggest two examples: constellations and anamorphosis. Let us start with constellations. When we look at a starry sky, we detect shapes that the stars form. It is hard to look at the relevant part of northern celestial hemisphere without noticing Corona Borealis for instance, composed of eight main stars forming what looks like an arc of a circle.

Now, although these eight stars appear to us as aligned, they are distant from earth in extremely different ways, roughly from 75 light years to 473 light years. From virtually every other perspective on these eight stars than from earth, nothing like an arc of a circle appears. And the astronomer has no use for constellations, it just does not make sense to their activity.

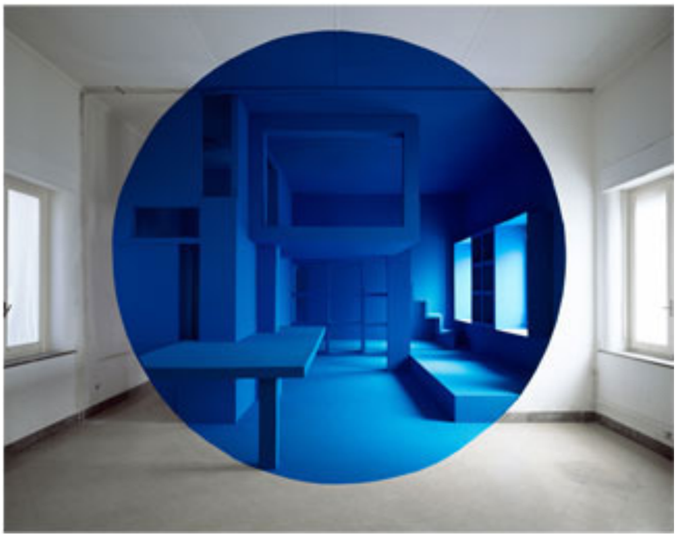
Shall we conclude from these facts that constellations do not exist? Surely not, as everything they are made of (stars) exist. But even thought the concept of CONSTELLATION do not make sense independent of us perceiving from earth the stars forming it, it supervenes on the response of our cognitive system to it (in this vein, cf. Dennett 1991).

Just because we cannot easily secure the reference of CONSTELLATION independent of a perspective does not mean that "Corona Borealis" fails to refer, or that perception of it is illusory. "Corona borealis" just refers to that set of stars that appear to well-seeing humans from earth as an arc of a circle, and that description just happens to refer to a real set of stars that the astrophysicist, or anyone from a distant planet, has no point in putting together.

Still, the constellation exists for everybody, for the astrophysicist as well as for the perceiver from a distant planet, as the constellation just is, by definition, that set of stars. Note that in that case, unlike maybe in the case of colors, it is perfectly possible (although pragmatically irrelevant) to identify the relevant set of stars independently of any "response".

My second example of an objective, although not mind-independent, feature of the world is that of anamorphosis. The French artist and photographer Georges Rousse makes a great use of that perceptual phenomenon. In one of his photographs for instance (see figure 4), one sees a blue disc superimposed on a white room full of white pillars and white surfaces:

FIGURE 4. A PHOTOGRAPH BY GEORGES ROUSSE



In fact, every bit of blue is just a bit of paint on the walls, pillars and other surfaces in the room. From the point where the camera lens was standing when the picture was taken though, an image of a translucid blue disc is reconstituted, as if it was a floating before the room. From any other angle, we would just see a room painted in blue in some spots and in white in other spots.

Now, we need not conclude from this apparent illusion that the disc in George Rousse's picture does not exist. It is just that it makes sense to conceive of what exists as a "blue disc" only from a very limited perspective. The pigments on the walls do cause our perceptual system to react in a certain way (perceiving a disc). Even if it cannot be relevantly conceived of as a disc from any other perspective, the referent of "that blue disc" is not necessarily empty, strictly speaking. There *is* an objective feature of the world referred to by "that blue disc" which just isn't mind-independent[105]. The dispositional, response-dependent account of color concepts elaborates on that idea.

In addition to being consistent with color science, and preserving the objectivity and truth-evaluability of color talk and thought, the response-dependence view has the advantage of accounting for the intuitive fact that we need color experience to understand a color concept. Without the appropriate cognitive system that confers (through similarity judgments based on the similarity of the responses) unity to the underlying diversity, it is unclear that we can

[105] Although one could argue it is best characterized as an illusion.

fully grasp the concept (though we may be competent with the term "red" – see below for more on this issue).

7.1.5. RD Properties and RD Concepts

I want to stress here that I will treat response-dependence as a property of concepts rather than as a property of properties. Both conceptions of response-dependence are possible. We could hold that response-dependence is a property of properties that are grounded in relations between objects and human subjectivity. Under that conception, an object would bear a response-dependent property in virtue of being such as to elicit a mental response from a subject under certain specified conditions.

But I prefer to remain as neutral as possible on metaphysics, not to impose unnecessary constraints on the ontology. Here are some reasons for taking response-dependence as a feature of concepts instead[106].

Take RED again. Intuitively, the different objects that are red do not have much in common apart from the fact that human perceivers detect in them a common property. This means that the objects that are red are more easily put together from an anthropocentric perspective

[106] Note that the notion of response-dependence is discussed in several other philosophical contexts. It is important in epistemology for questions relative to the distinction between basic and derived knowledge. Knowledge acquired by definitions would be semantically derived and knowledge acquired by ostension would be semantically basic. As some concepts must be semantically primitive in order for the whole conceptual system to function (Wittgenstein et al., 1969), and some have suggested that semantically basic terms are necessarily response-dependent (Jackson and Pettit, 2002). For instance, to grasp the concept RED in a semantically basic way (not in the way a color blind person would master the term), one needs to have color experiences, whatever that is. Some have suggested that this basic fact would confer *immunity to error through misidentification* (IEM) to response-dependent concepts (Holton, 1991).

than from a purely cosmocentric perspective. From a purely cosmocentric perspective, there is not much in common between a red tomato and an English phone booth. This is a first motivation to focus on the concept RED rather than on the red property, because the red property itself is interesting only inasmuch as humans perceive and conceptualize it.

Moreover, we can hardly know anything about properties in general. Arguably, we might not be even *able* to know anything about them, as neo-Kantians think (see for instance Langton 1998)[107]. Why then ascribe properties to properties when that is not necessary?

What we know, after all, is not really that red things are things that look red to normal observers, but rather that we humans who perceive, think and talk about red seem to apply the notion to those things that look red: that is a psychological rather than a metaphysical observation. A response-dependent view of redness, then, would not be the view that red things are disposed to look red, but rather the view that (at least some level of) people's representations of red things reflect such a belief.

Second, and this follows from the previous point, I take it that human categorization behavior is a central explanandum in philosophical investigation.

Finally, since I want to invoke the notion of response-dependence in order to provide a plausible model of slurring concepts, I shall not talk of response-dependent *properties* because that would presuppose that there *are* such properties as being a "kike" or being a "boche".

But at this stage we want to keep open the possibility that these concepts have an empty extension (so as to account for their defectiveness for instance). And only response-dependent *concepts* - as opposed to response-dependent *properties* - are compatible with the empty-extensionality thesis.

Color concepts could be non-empty response-dependent concepts - they would successfully apply to whatever properties actually triggered the relevant perceptual responses -, whereas

[107] Or see Fodor's (1981) discussing Kant: "You must not think that because there are chairs and horses and sensations in our representation, that there are correspondingly noumenal chairs and noumenal horses and noumenal sensations. There is not even a one-to-one correspondence between things-for-us and things in themselves." (p. 63)

SCs could be empty response-dependent concepts - if they are actually brought about not by a response to actual objects and properties but by a response to illusions or imaginary objects. SCs would be somehow recycling a preexisting psychological mechanism of response-dependence, but that shall not entail that their functioning - especially regarding potential emptiness - is the same in every respect.

For these reasons, I will from now on put aside metaphysical talk of "response-dependent properties" and rather focus on psychology, that is on "response-dependent concepts". Keep in mind that after all that what we want is an account of slurring thought and talk.

7.1.6. Response-Dependence and Relativism

Response-dependence is sometimes seen as a first step towards relativism. As red is linked to looking red to subjects, it could seem superficially that red in fact reduces to red-for-x or red-for y. In a world where a class of perceivers is sensitive to red, things would *be* red, but in a world without such subjects, like in a counterfactual world where humans have all evolved to be color-blind, the objects would not count as red. And in a world where things that are green for us today happen to trigger an experience or red in normal conditions, then green things would be red.

Hence a moral response-dependent theorist could promptly conclude that nothing is right nor wrong *per se*, as RIGHT and WRONG are dependent on human responses and human responses are variable. This would give rise to a form of moral relativism. Koons (Koons 2003) for instance argues against response-dependent accounts of morality on the basis that the kind of relativism it gives rise to contradicts the universality and objectivity of moral judgments.

The worry is understandable, but I argue it is misplaced. One can tie morality to human affective states without thereby endorsing the conclusion that in a carefully designed counterfactual situation, wrong things would be good. All the theorists need to do is to rigidify. As Vallentyne (1996) puts it,

> The key issue is whether, for a given claim of wrongness, the relevant responsive dispositions of the beings B and the conditions C are the same no matter what the time, world, and agent of the action being assessed are. If the relevant responsive dispositions and conditions are fixed and the same for all actions evaluated, the account is rigid, and if not, the account is non-rigid. (pp. 103-104)

When we say that red things are just the things that look red to normal observers in normal conditions, we do not say that anything that would look red to normal observers in normal conditions in other worlds would be red.

The reference of RED is fixed by the *actual* observers and conditions of observation. If this was not the case, that is, if RED happened to refer to the objective physical and chemical properties X, Y and Z of surfaces (because X, Y and Z would be the only properties that the human perceptual system responds with a red* response to), then RED would simply have objects with properties X, Y and Z as its extension. Thus, in a world where X, Y and Z would trigger a blue* response in subjects, the objects with X, Y or Z would not be simple blue objects with a red appearance; they would count as plain red objects.

This does not align with intuition: if there is another planet whit an atmosphere such that my red tomato appears to be blue when I travel there, I would not count my tomato to *be* blue, but rather to *appear* blue while being in fact red.

In other terms, the observers and conditions of observation can legitimately be rigidified. What one is trying to do in characterizing redness is to have a characterization of an aspect of *our* world, not of other possible worlds. "What are the things that are red?" is an empirical question, it does not apply to other possible worlds where objects appear differently.

The answer that a response-dependentist could give to that question is that the things that are red are those things that cause such and such reactions in such and such cognitive systems. Once we know which are those things that are red, we can in principle characterize them independently, without having recourse to the responses of the cognitive system.

We need human responses to know what are the things that are red, but once we know it, we do not need human responses anymore. Only once the reference of RED is fixed can one extend the question "what are the red things?" to other possible worlds. In all other possible worlds, whatever the human responses are, red things are the things that have the property of redness - defined in the actual world - such as X, Y and Z.

The question arises of why we say then that RED is response-dependent if it simply refers to objects with properties X, Y or Z. A simple answer would be that it is response-dependent because it is only a posteriori that we know it refers to objects that have properties X, Y or Z, whereas it is a priori that it refers to those objects that look red to normal observers in normal conditions. I come back to the a prioricity of response-dependence below.

I now turn to an examination of the hypothesis that slurring concepts are response-dependent concepts, which will require a more fine-grained discussion of response-dependence in general.

> ...most things that interest us in our normal human lives are response-dependent, goodness-wickedness, beauty-ugliness, attractiveness-repulsiveness, being humanly meaningful vs. being meaningless and empty. In contrast, most things that are metaphysically important are not response-dependent. [...] response-dependence belongs to the manifest picture we care about humanly, independence to the deep reality we care about scientifically (Miscevic 2011b, p. 80)

In order to have a better grasp on the precise notion that will be of interest for slurring concepts, I will prepare the discussion with a few necessary distinctions and clarifications. I start with large-scale distinctions and continue with more and more fine-grained notions, gradually narrowing the focus until we encounter the notion of response-dependence that will enlighten the question of slurs.

These clarifications will also be useful to a better understanding of dispositional concepts in general and will provide a general map for the discussion**.**

7.2.1. Dispositional Effects

I have evoked the notion of a *disposition*. Let me now introduce a little bit of terminology clarifying the surroundings of the notion. First, there is what makes distinct dispositions distinct. For instance, it is a disposition that makes things "fragile" and it is a disposition too that makes some things "ephemeral". As a rough approximation, an object is fragile if it has a disposition to break under certain pressure (as we saw above), and an object is ephemeral if it has a disposition to disappear relatively rapidly. Hence, although "fragile" and "ephemeral" both target a disposition, they are not the same disposition.

They are not the same disposition because of what they are disposing the objects to undergo. What distinguishes the predicate "fragile" from the predicate "ephemeral" is precisely what the disposition is disposing the objects to. I call that the *dispositional effect* of the disposition. The relevant disposition of a fragile bottle has the dispositional effect to make the bottle break under pressure, and the relevant disposition of an ephemeral insect has the dispositional effect to make the insect disappear relatively rapidly.

Second, the actual set of physical properties responsible for the bottle's disposition to break is the *realizer* of the disposition. Figure 5 and 6 illustrates the relation between these three notions:

FIGURE 5. FROM THE REALIZER TO THE DISPOSITIONAL EFFECT

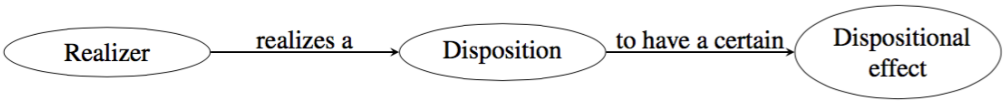
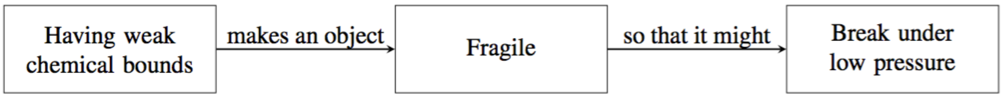


FIGURE 6. EXAMPLE OF FRAGILITY



7.2.2. Dispositional Concepts

Now I turn to a brief exploration of different families of dispositional concepts, which will be especially useful in the case of SCs which involve subjective dispositional effects. Note first that it is taken for granted that the dispositional effects we are interested in are *constitutive* of the disposition rather than *parasitic* on it (Wedgwood, 1998).

Consider a "stable" object. As stable, the object has the disposition to stand still under certain turbulences (approximately). But stable objects might also happen to have the disposition to elicit such and such behaviors in ants walking on it. For instance, ants might walk in straight lines on stable objects more often that on non-stable objects. What makes the object stable is its standing still under turbulence, not its giving rise to certain behaviors in ants, even though it is precisely its standing still which elicits ant's specific behavior. The object's being stable has the constitutive dispositional effect to stand still under certain turbulences, and the *parasitic* dispositional effect to elicit such and such behaviors in ants.

There is an asymmetrical dependence between these two types of dispositional effect, the constitutive and the parasitic: there could not be such and such ant's behavior without the object standing still, but the object could well be standing still without the ant's walking on it.

Among the dispositional effects that are constitutive, we can draw a line between effects that are *monadic* and those that are *relational* (effects on other things). For example, a product being "volatile" consists in its having, as an approximation, the objective disposition to evaporate easily. That is a monadic effect, because the disposition to evaporate does not involve any other object than itself.

On the other side, an object being "stimulant" consists in its having the disposition to speed up the heart-beat of certain creatures with a heart. That is a relational effect, because the effect is an effect on another object (here hearts).

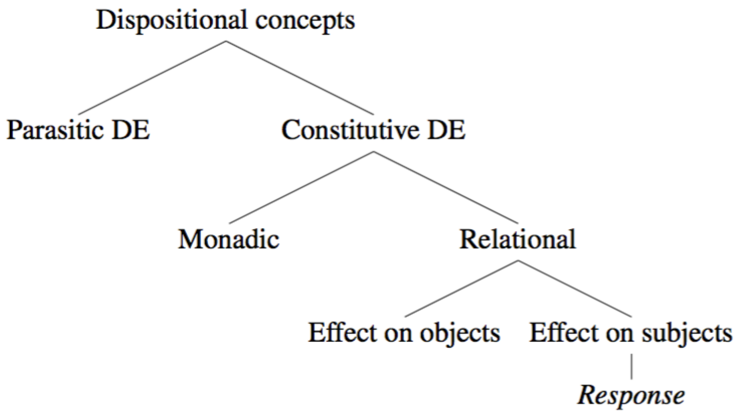
Furthermore, among the relational effects, one may distinguish the effects induced on (possibly inanimate) *objects* and the effects induced in *subjects* (and especially humans – I will ignore other subjects in what follows). The effect of stimulants belongs to the second category. In this type of case the dispositional effect may be called a *response*.[108]

We shall focus on the cases in which the dispositional effect is a response induced in human subjects, that I call *responders*[109]. Figure 7 illustrates these distinctions among dispositional concepts, and locates response-dependent concepts at the bottom:

[108] Note that in the way I use the term, a "response" does not have to be of a sensory character to qualify the concept as response-dependent.

[109] Note that the responders need not be identical to the possessors. Martians, who (let us assume) do not have hearts, may still possess, master and use the concept of a "stimulant" which expresses the disposition to elicit a certain response in another group of subjects, namely creatures with a heart.

FIGURE 7. SITUATING RESPONSE-DEPENDENT CONCEPTS AMONG DISPOSITIONAL CONCEPTS



In a nutshell, response-dependent concepts are concepts whose extension is determined through human subjectivity, that is by the cognitive responses of (a class of) subjects. A concept is response-dependent when it targets a dispositional property of an object to elicit a mental response from an agent under specified conditions.

Typically, secondary quality concepts such as RED or HOT are response-dependent. It is canonic since Johnston's "basic equation" to characterize response-dependence with one version or another of a biconditional (Johnston et al., 1989). I give below a schematic version of such a biconditional, where x stands for an object, F for a property (e. g. "red" or "circle"), S for a class of subjects (e. g. "healthy human beings" or "perfect square detectors"), R for a class of cognitive responses (e. g. "activation of neural network n" or "sensation of pain") , and C for a class of conditions (e. g. "daylight" or maybe "ideal circumstances"):

RDB: x is F iff x is disposed to trigger response R in subjects S under conditions C.

There are three relevant elements in the right-hand side of the biconditional. That is, x's being an F or not crucially depends on the encounter of three things.

The first is the cognitive response. It is necessary for x to be an F that x provokes a certain kind of cognitive response. Second, the response must be a response of certain specific kind of subjects. The class of relevant subjects could be all human beings in some cases, or maybe a single individual in other cases, or anything in between. But what is crucial is that a class of subjects, standardly referred to as the "judges", is specified.

Finally, some external conditions must be specified. It cannot be the case that x is an F if it triggers the right response from the right subjects in any sort of conceivable conditions, for arguably there always are conceivable conditions so improbable and unnatural that you can make the right subjects have the right sort of response in it. For instance, in a situation where the daylight is red, normal subjects would perceive a white wall as red. If the concept F is response-dependent, then for an object to be F it must be disposed to trigger the right sort of response in the right kind of subjects under the *right sort of conditions*.

Note that we can conceive of a dispositional property that would fail to stand in the right relation to the subject's reactions, even though it should intuitively fall under the response-dependent biconditional. We can for instance imagine a cleverly designed object that is red, that is, it has all the microstructural properties X, Y and Z of redness so that it is disposed to appear red, but that on top of these properties, the object has the strange habit to become invisible under normal lighting.

Such an object would have the right dispositional property to qualify as red, even though it is not actually the case that it would appear red to normal observers under normal conditions. It would thus not have the disposition itself, but rather

> what Ian Hunt once called (such a disposition) a "finkish" disposition, one that would vanish if put to the test. (Lewis 1989, footnote 6 p. 117).

Another example is Johnston's (1992) shy chameleon, who is green but blushes and becomes red instantaneously when put into viewing conditions. It is clear in that case that the simple fact that the chameleon is shy shall not make it lack its real color (green).

This means that response-dependent biconditional should be understood abstractly and disconnected from the other (possibly interfering) properties that the relevant objects might or not have (such as shyness for the green chameleon)[110]. This is also the reason why "ideal" conditions are so often invoked when talking about response-dependence.

[110] This is of course not sufficient to address the various sorts of problems raised by finkish dispositions, by we can leave these sorts of issues aside.

7.2.3. Global Response-Dependence

Arguably most, if not all, concepts satisfy some unconstrained version of RDB. Indeed, for any F, it is arguable that "x is F iff x is disposed to trigger in subjects the application of their F-concepts to x in conditions in which subjects are infallible in their F-applications".

Take the concept of a RECTANGLE for instance, which seems to be perfectly objective and independent from subjectivity. Rectangular things might well provoke a distinctive kind of cognitive response in human subjects under some circumstances (Vallentyne, 1996). We could then construct the following characterizing biconditional:

(167) x is *rectangular* iff x is disposed to cause (perfect) rectangle detectors to have a perception of a rectangle under (perfect) circumstances for geometrical perception.

Indeed, isn't it trivially true that anything disposed to cause a perfect rectangle detector to detect a rectangle is a rectangle? Isn't RECTANGLE response-dependent then, as it is governed by a biconditional like (167)? Fodor (1998) has a similar remark on everyday concepts such as DOORKNOB, which looks a lot like a global response-dependence view of natural language concepts:

> My story says that what doorknobs have in common *qua* doorknobs is being the kind of thing that our kind of minds (do or would) lock to from experience with instances of the doorknob stereotype. (Fodor 1998, p. 137)

Similarly, take the concept BIRD. Arguably, it is true that:

(168) x is a *bird* iff x would cause humans to have an avian experience in circumstances that are favorable for doing so.

As a consequence, a response-dependent view of slurring concepts should be read as claiming that SCs are *distinctively* or characteristically response-dependent. Hence, proponents of RDA should fill out the canonical RDB-schema in sufficient detail to give a substantial theory that is specific to SCs, just like response-dependent accounts of color concepts should be substantial enough to distinctively characterize color concepts. Here are a few general remarks that will help to address that worry, in specifying more carefully a substantial conception of response-dependent concepts.

First, Wright (1988) notes the importance of specifying explicitly the conditions C in the biconditional instead of using an empty placeholder like "favorable conditions", because that would make the basic equation trivially true and hence uninterestingly generalizable.

Second, as in any biconditionals and definitions, we should also be careful about the order of determination. That a red object is disposed to look red (to normal observers in normal conditions) is expected and less surprising than the fact that looking red (to normal observers in normal conditions) could be sufficient for an object to *be* red. Redness is surely linked to a disposition to look red, but is the tomato red because it is disposed to look red or is it disposed to look red because it is red? Most response-dependists argue for the former, that is, they *define* "being red" through "looking red", and not the other way around.

Now, it is true that some things look square simply because they are square, and it might be that all things that are square are disposed to look square to normal observers in normal conditions. But is it their looking square that make them square? If not, squareness is not really response-dependent in the sense that the biconditional would not have the proper order of determination. This observation is connected to the *essential* role that the response must play for a concept to be truly response-dependent, a role I evoke in the following remarks.

Third, the possibility that most concepts may be connected to certain cognitive responses does not undermine the project: I focus on these concepts that are essentially connected to responses.

In the case of bird, the avian experience does not guide categorization. Where the concept "red" is essentially the concept of what causes red experience, the concept "bird" is not the concept of what causes avian experience, even though birds do cause avian experiences. Response-dependent concepts are essentially connected to certain responses, and a way to capture this intuition is to embed the basic equation under a modal operator:

(169) Necessarily, x is F iff x would cause humans to have a F* experience in conditions C.

or perhaps:

(170) It is constitutive of F that it applies to x iff x would cause humans to have a F* experience in conditions C.

It now becomes appealing to apply the response-dependent model to SCs, because what intuitively distinguishes SCs from NCs is the presence of a specific affective or psychological response (emotional/social etc.).

In earlier work with Recanati (unpublished master thesis), I started to examine the view that STs express response-dependent concepts such as RED, that is, concepts such that their possessors are the responders. The intuition is that

> …to use a pejorative term, perhaps a certain relation must hold between the speaker and the referent, namely the possession of the attitude of contempt. (Saka 2007, p. 128)

Intuitively, the response, which involves negative valence, is normally present when KIKE is deployed, but not when JEW is. Making this response constitutive of SCs seems a good start to account for this characteristic hotness/expressivity. We can then start considering RDA:

> **RDA**: SCs are response-dependent concepts

As the most paradigmatic cases of response-dependent concepts are color concepts, we will first try to propose an account of SCs on the model of RED.

We saw that a paradigmatic class of concepts for which Response-Dependent Accounts (RDA) have been popular are color concepts like RED. Hence, we start by investigating the plausibility of a RDA account of SCs, according to which SCs are patterned rather closely on concepts like RED. Assume the following RDB governing RED:

> x is *red* iff x would cause human beings with a well-functioning perceptual system to have a red* experience under standard lighting conditions.

That is, roughly an object being red consists in its looking red to normal observers in normal conditions. Note that "red" is the name of the objective property and that "red*" is the name of the subjective experience. The purpose of having "red*" and not "red" on the right hand side of the biconditional is to refer unambiguously to a property of the experience, thereby preventing the formula from being circular.

But if the RED bi-conditional holds, does it follow that subjects know *a priori* that red is a secondary quality? If it is the case that the use of "red" is governed by the RDB, then users of "red" must master the RDB. In the same way as we know *a priori* by that bachelors are unmarried man, it is now supposed to be known *a priori* that red things are what provoke a red sensation[111].

This seems false. It would be odd to suppose that normal possessors of RED know that RED is a secondary quality. It even took centuries of color science to start casting doubt on the idea that color was akin to shape. Application of the concept RED seems to work in a way simpler manner: subjects perceive a property (redness) in an object and apply RED to it, just like for shape and movement.

Unlike other cases like maybe COMFORTABLE, subjects need not know that RED is response-dependent, that is, that it applies to a dispositional property to trigger a certain kind of effect in them (or in other subjects) under the right sort of circumstances. The mastery of COMFORTABLE more plausibly involves the knowledge that the targeted property is dispositional in this sense. When we apply the concept COMFORTABLE to a couch for instance, we do so because we suppose that the couch would be pleasant if we were to sit or lie on it (or something along these lines).

We do not directly, perceptually detect a property and apply the corresponding concept. In the case of "comfortable", knowledge of the biconditional is part and parcel of the mastery,

[111] I distinguish this issue from the question of whether it is knowable *a priori*, by conceptual analysis, that RED is governed by a RDB. It is for instance conceivable that subjects do not know that RED is governed by a RDB even though it is *a priori* knowable by astute conceptual analysis, a bit like the causal theory of reference is arguably knowable a priori even tough it was ignored before Kripke. My point here is simply that the layman does not know a priori that RED is governed by a RDB? I leave aside the question of whether is is knowable a priori.

hence the possession, of the concept. That is a case of what I call a "transparent" response-dependent concept.

But RED is not like COMFORTABLE. Normal naive subjects seem to apply RED as if it was a concept of a primary quality. As Pettit (2003) puts it,

> While redness does not become salient to us as something that plays a certain dispositional role – as something that has the observed effect of making things look red – it does become salient in virtue of playing that role, actually making things look red to us. (p. 225)

Although RED is governed by a RDB, in applying RED, we are guided by our red* response to the objects we encounter. The biconditional governs one's application of the concept, but from the subject's point of view, it is as if she was detecting the property in the object itself. Miscevic (2011b) makes a similar point:

> ...a tomato's being, say, red in a scientific sense, is being such as to cause in normal observers under normal circumstances the response as of seeing phenomenal-red, experienced intentionally as a simple property of the surface of the tomato. (p. 77)

Schematically, we call red these things that happen to provoke a perception of red in us. The objective properties provoking the perceptual response may be unknown to us, whenever we have the response on which we ground RED, we are entitled to apply RED; that is, precisely because RED is a concept one usually grounds on red*, RED picks out whatever complex and diverse properties are responsible for red*. As Jackson (1998) puts it, most of us see colors

> as properties of things and perhaps independently describable properties of things, that are unified and important only in virtue of their association with our color sensations. (Jackson 1998, pp. 244).

But if subjects do not know the RDB, how can an RDB still govern the application of the concept without its possessors knowing that is the case? That is because, as Pettit and Jackson notice, there is room for another way in which a concept may be governed by the biconditional, than the possessors of the concept knowing that the biconditional holds (Pettit, 1991, Jackson and Pettit, 2002).

Suppose for example that the responders are the possessors of the concept. Then they can let the response guide them in their application of the concept: they blindly apply the concept to whatever triggers the response in them. They use the response itself as a reliable indicator that the concept is applicable. In this case (opaque response-dependent concepts), the biconditional governs the application of the concept but does not have to be known in order to possess the concept.

Color concepts such as RED seem to belong to that category. In these cases of opaque response-dependent concepts, the possessors being themselves responders can fully rely on their response to apply the concept. That is a case of what I will call "opacity", as opposed to "transparency":

> **Opacity**: A response-dependent concept is *opaque* when knowledge of the governing RDB is not necessary for subjects to possess the concept. The RDB is thus not *a priori*.

I argue that RED is opaque. Most of us do not even know that RED is governed by a RDB, but still, we correctly apply the concept to red objects. We let the response guide our application of the concept, and whether or not we are aware that the concept in fact applies to the objects which trigger in a certain class of subjects - including ourselves - a certain kind of perceptual response in certain circumstances, is irrelevant.

This opacity accounts for the difference between our application of RED and that of COMFORTABLE. If it seems that we detect a simple objective, non-dispositional property of redness in objects, it is not that redness is a primary quality, but rather that being the relevant sort of subjects in the relevant sort of conditions ourselves, we need not know that red is a secondary quality to deploy and apply RED correctly. All we need to do is to apply the concept whenever we believe we are in normal conditions and feel the response.

Since we do the same for COMFORTABLE (we apply COMFORTABLE when we detect comfortable objects), there is no phenomenological difference between such concepts and opaque concepts for secondary qualities such as RED. But the absence of phenomenological difference shall not lead to a theoretical identification of the two kinds of concepts, because there are other criteria than phenomenology to distinguish between two kinds of concepts, such as the meta-semantics distinction between transparency and opacity.

Now of course, there are instances of applications of RED by subjects who do not seem to be responders. Color-blind subjects for instance do not have the red* response, but they still use the term "red" and have mental analogs of it. An omniscient color-blind color-scientist might even be equally competent to everyday perceivers in their categorization behavior. That RED is opaque means that knowledge of the RDB is not necessary for concept possession, but is it sufficient?

Do these subjects who apply RED without being responders and simply in virtue of their knowledge of the RDB possess the concept RED? The question is crucial for SCs, as many of the properties that RDA could account for, such as hotness and projection, will depend on whether or not the speakers/thinkers are taken to have the response. And since we intend to apply the RED-model to SCs, it is important to take a stance on issues regarding the possession and individuation of response-dependent concepts.

There are two intuitive theoretical options to account for the use of RED by color-blind people: either they possess the same concept, in which case the response red* is not constitutive of the concept, or they use it in a deferential manner and hence possess a different concept of RED.

The choice between these two options depends partly on what requirements we impose on subjects' cognition and behavior for them to count as possessing a concept. I will call the two main positions regarding this issue the "Reflexivity Thesis" and "Non-Reflexivity Thesis". Let me first introduce the notion of reflexivity.

> **Reflexivity**: A response-dependent concept is *reflexive* when its possessors possess the concept in virtue of being responders.

Reflexivity imposes a constraint on concept possession: in order to possess the concept, subjects must have the response. When a response-dependent concept is opaque, the response is sufficient for concept possession; when it is reflexive, the response is necessary for concept possession.

Note that reflexivity asymmetrically entails opacity. If a concept is reflexive, then, by definition, its possession requires having the response. We can thus not possess the concept unless we are responders. Having the response is therefore a necessary condition for the normal possession of a reflexive concept. But as soon as we are responders, we can let the response govern the application of the concept. Whether or not we have knowledge, on top of that, of the governing RDB becomes irrelevant. Knowledge of the RDB is thus unnecessary as soon as we have the response. And the fact that we can possess a response-dependent concept without necessarily having to know the RDB is what I coined "opacity" above. Hence reflexivity entails opacity.

The converse is not true though. Logically, a response-dependent concept can be opaque and non-reflexive. That would be the case of a concept which would be possessed both by i) responders who do not know the RDB but can possess the concept because it is opaque, and by ii) subjects who know the RDB but are not responders themselves. So opacity does not entail reflexivity.

We agreed that RED was opaque, but before applying the RED model to SCs, we must ask ourselves: is RED reflexive? Answering this question trickier, because it hinges on questions about concept individuation. It amounts to wondering whether someone can possess the concept without having the response at all, simply in virtue of knowing the RDB. Does our color-blind omniscient color-scientist possess RED for instance? If yes, then the concept is not reflexive. If not, it is reflexive.

One reason to lean toward the reflexivity of response-dependent concepts like RED has to do with the reason response-dependent concepts were introduced in the first place. Theorists had the need to introduce response-dependent concepts in order to account for the behavior of some secondary quality concepts which seemed connected to non-conceptual responses in a way that other concepts were not.

If the possession of RED was possible in the absence of a response, in what sense would RED be different from COMFORTABLE? In what crucial sense would it depend on a non-conceptual perceptual response?

Response-dependent concepts must form a distinctive class of concepts, and their specificity is precisely the strong connection they entertain with a certain non-conceptual response. For the response to truly characterize response-dependent concepts, it cannot be merely associated to it.

If anyone, including non-responders, can possess the concept, then we loose what is characteristic of these concepts as opposed to other concepts. We would have a concept of RED that some possess in virtue of the red* response and that others possess in virtue of knowledge of the RDB. But both instances of RED would be instances of one and the same concept. This is not compatible with the attempt to account for secondary quality concepts as a *distinctive* class of concepts grounded on a cognitive response.

To have truly response-dependent concepts, the concept must be constituted by the response. This is why I will consider that RED is opaque and reflexive: that is, the response will be understood as constitutive of the concept in the sense that having it is a necessary condition for concept possession. The concept that color-blind people may possess and ground on the RDB would, under that view, not be instances of the concept RED.

So I will now apply the RED model to slurring concepts, under the assumption that RED is a response-dependent, opaque and reflexive concept.

7.3.3. Application of the RED Model

The evolution from color error theories to theories of response-dependence that we presented earlier has had a parallel, to a lesser extent, in debates concerning value. In metaethics error theories are often applied to moral concepts (Blackburn, 1985, Sayre-McCord, 2014). Moral error theories claim that moral judgments don't succeed in ascribing properties to events or actions. "Stealing is wrong" is thus not true, because nothing in the world instantiates the property of wrongness.

Consider this observational moral judgment: someone sees a man beating a dog and thinks: "That is bad"/"That is wrong". In this example, there is an objective event in the world (a man beating a dog), and an observer perceiving this event and forming a thought about it.

An error-theorist will analyze the thought - and the sentence used to express it - by saying that the objective perceived event provokes certain beliefs and emotions, which are then "projected" onto the event (Hume, 1739). Moral error-theories thus imply that we falsely believe moral properties are in the world, just like color error-theories imply that we falsely believe colors are in objects.

Like for colors, response-dependent accounts of value are starting to develop in place of moral error theories (see e. g. Johnston, Smith and Lewis 1989 or Miscevic 2006). Restricting the investigation to SCs, I will now evaluate the extent to which one can account for the specificities of SCs in arguing that they are response-dependent concepts, on the model of RED. It is tempting to model SCs on color concepts.

Indeed, as the problem of SCs is the question of how a seemingly descriptive and seemingly expressive components relate, and since secondary quality concepts are also tightly connected to certain non-conceptual states, pursuing the analogy could be promising. How do emotions intervene in (moral) categorization? A bit like color percepts intervenes in color categorization.

Consider RDAred, a first attempt at applying the RED model to SCs.

> **RDAred1**: x is a *SC* iff x would cause [subjects S] to have [response R] under [conditions C].

RDAred1 is a schema of a RDB applied to SCs in general, with placeholders for S, R and C. But for each particular SC, we need to specify a relevant groups of subjects, a kind of response, and of conditions. Ideally, with enough commonalities between judges, responses, and conditions, we could give substance to the placeholders and put forward a generalized version of an RDAred applying to all SCs.

I will now briefly elaborate on each of the three dimensions (for subjects, responses and conditions respectively), so as to gradually reach a more substantial version of an RDAred.

Let us start with the group of subjects. Who are the judges for a given SC? Whose cognitive response is relevant for the determination of SC's target? And more generally, is there anything common between the different groups of judges for different SCs?

Intuitively, the most likely subjects whose responses are relevant for e. g. BOCHE are precisely the people who are prejudiced against German people, in short germanophobes. Similarly, who else than the anti-Semites will be responsive in the right way to the target of KIKE? It is tempting to conclude, more generally, that the relevant group of subjects in the RDB for SCs are the subjects who are prejudiced against the targets.

If RED is grounded on the red* perception of normal human subjects, and if SCs are to be modeled on RED, then SCs must be grounded on the responses of normal prejudiced subjects: germanophobes for BOCHE, anti-semites for KIKE, and so on and so forth. Let us try:

> **RDAred2**: x is a *SC* iff x would cause racists to have [response R] under [conditions C].

But it is immediately clear that the group of judges should be further specified, as *i)* not every SC is a case of "racism" (e. g. sexist SCs), and more problematically, *ii)* the set of all racists is too broad a group to differentiate between different SCs. More, it is not so clear that there is such a thing as a group of "racists".

For instance, male chauvinists are as irrelevant to the extension of KIKE as anti-Semites are irrelevant to the extension of BITCH. We shall thus individuate the relevant group of subjects in a more fine-grained manner. As a general rule, it seems that the group of subjects whose response will be relevant for the extension of a SC is the group of people who despise/hate/are prejudiced against the SC's target.

How can we individuate these groups for each SC? We could use NCs as a heuristic: since SCs and NCs share their target (by definition), we can refer to the relevant group of prejudiced people in a non-circular way as "NC-phobic people". That is, for KIKE, the group of NC-phobic people will be those who have the relevant sort of negative response to Jewish people (the anti-Semites), and for BITCH, it will be those who have the same kind of negative response to women (e. g. the male chauvinists).

We can thus construct a more satisfactory version of RDAred:

> **RDAred3**: x is a *SC* iff x would cause NC-phobic people to have [response R] under [conditions C].

Let us now turn to the nature of the response R. In the case of RED, the response was a perception of red that we named red*. Red* could either be a phenomenological experience of red perception, or a specific sort of event in the brain of subjects. Response-dependent accounts of concepts need not take a stance on such issues in the philosophy of mind.

But in the case of SCs, we cannot sensibly say that the response is perceptual like in the case of red*. First, many possessors of SCs have never even been acquainted with their targets. There exist homophobes who have never met a homosexual person for instance. The idea itself repels the homophobes, more than the individuals who they might have seen or talked to. But since perception requires acquaintance, if there are cases of homophobes deploying FAGGOT without having had acquaintance with homosexuals, the response cannot be perceptual.

Second, even though SC possessors have sometimes been acquainted with their targets, there are many SCs that target groups whose membership is not accessible through a perceptual medium. All nationalities are like that. What makes someone German, for instance, is her citizenship. But citizenship is not visible on someone's physical characteristics (even though some visible characteristics correlate with the possession of a passport, such as the accent, behavior in certain situations etc.)

So even though germanophobes are perceptually acquainted with Germans, their response is not triggered through perception. The same holds for many SCs that can be applied without any particular physical characteristic other than group membership being instantiated: FAGGOT, KIKE, WOP etc. So if the response is not of the perceptual sort, what does it consist in?

Moral emotions are promising candidates. It is clear that racism and prejudice involve some sort of a negative emotion or another. Germanophobes dislike Germans, or hate them, or something along those lines. An application of that possibility to RDA gives us:

> **RDAred4**: x is a *SC* iff x would cause NC-phobic people to have negative emotions under [conditions C].

But "negative emotions" is too broad, we need to specify it because not all negative emotions felt toward a target are instances of racism. It could be an emotion of reject or disgust. But again, it seems that disgust and reject are not social enough as emotions to count as the canonical racist response. It rather seems that the negative emotion should be a *social* emotion, on par with shame, guilt, or embarrassment for instance.

Contempt is an interesting candidate (cf. Jeshion 2013a). Contempt is indeed a negative emotion that is at the same time a moral emotion and a social emotion, which seems well fit for the kind of racist response we are looking for. And probably all subjects who possess an SC share a kind of contemptuous feel toward their targets. If contempt is the relevant kind of response, our schema of a SC biconditional would become:

> **RDAred5**: x is a *SC* iff x would cause NC-phobic people to experience contempt under [conditions C].

Now, even though contempt could be a necessary ingredient in the response, it might not be sufficient. There are indeed many reasons for contempt other than prejudice and racism. The simple fact that someone is regarded contemptuously does not make it a case of slurring thought: an anti-Semite could for instance feel contempt for John without it being a case of anti-Semitism, and so even if John happened to be Jewish.

So the response is probably more fine-grained and complex than that. Hopefully we need not give a full specification of the response, even more so if it is complex hard to capture. There is no *a priori* reason to think that the characteristic kind of response involved in slurring thought is a simple natural kind. The typical response could at the same time be shared across all cases of racism and involve several dimensions: social, moral, perceptual, emotional etc.

Because it will be hard to identify a distinctive kind of response involving the many dimensions involved in prejudice and racism, I will instead start by using the place-holder "yuk*" to directly refer to the relevant response, whatever it turns out to be. We thus obtain:

> **RDAred6**: x is a *SC* iff x would cause NC-phobic people to have a yuk* experience under [conditions C].

And finally, what are the conditions under which NC-phobic people should experience the yuk* response for the concept to count as a SC? What are the "normal" conditions under which e. g. germanophobes feel yuk* toward Germans, hence grounding BOCHE?

One method to determine these conditions could be to wonder what are the "non-normal" conditions under which a Germanophobe would *not* detect her target. A case where a German pretends to be an Englishman, dressing and speaking as an Englishman, would not provoke the appropriate kind of germanophobic reaction onto Germanophobic people. The normal conditions might thus include some perceptual dimension such as a sum of apparent features that NC-phobic people are responsive to.

But arguably a strictly perceptual set of conditions will not do. There are cases of SCs targeting people who are not associated with any set of perceptible characteristics. Simply learning that a certain individual is an NC can be enough for the target to be counted as a SC.

Another possibility is that germanophobes have the response when thinking about their targets. We could thus invoke Lewis' (1989) conditions of "imaginative acquaintance" in his dispositional account of value. For instance, NC-phobic people would experience a yuk* response either when perceiving *or imagining* their targets:

> **RDAred7**: x is a *SC* iff x would cause NC-phobic people to have a yuk* response under conditions of imaginative acquaintance with x.

Imaginative acquaintance is a notion that should be clarified, and it is likely to be insufficient to fully account for the broad range of applications of SCs. The best might be to stick to a placeholder such as "normal conditions of slurring thought", as the details of the conditions will not play a role in a general discussion of response-dependent accounts of slurring concepts. Here is then our final version of RDAred, whose explanatory pros and cons we are now going to assess:

> **RDAred8**: x is a *SC* iff x would cause NC-phobic people to have a yuk* response under [normal conditions of slurring thought].

7.3.4. Reflexivity and Indexicality

The properties of reflexivity and that of indexicality are closely linked. Under another way to conceive of RDAred8, SCs can be understood as a species of indexical concepts. Saka noticed about ten years ago:

> So: is (1) [Nietzsche was a kraut] true, false, or neither? One possible view is that the truth-evaluable content of (1) is given by the sum of the cognitive contents (ψa) [S thinks that Nietzsche was German] and (ψb) [S disdains Germans as a class]. Since the content of (ψb) is affective and non-propositional, the cognitive content of the whole amounts to just (ψa); hence (ψ) is true. Another possible view identifies the truth of (1) with the correctness of (1') [As a member of the Anglophone community, S thinks 'Nietzsche was a kraut'] and the correctness of (1') with that of (ψa) & (ψb). In this case, some will hold that it is never correct for anyone to disdain Germans as a class and hence (1) is false. Others will hold that if S has personally suffered at the hands of genocidal Germans then prejudicial disdain on the part of S may be legitimate and therefore, in that sense, correct. On this view *the truth-value of (1) is indexical*. [my emphasis] (Saka 2007, p. 142)

In our terminology, as soon as we give an account of SCs modeled on RED as a reflexive response-dependent concept, it becomes possible to treat SCs as indexical concepts and STs as indexical expressions - in Kaplan's technical sense (Kaplan, 1979, 2001). Since the possessors of SCs are the responders, the content of STs, under that view, is the property referred to by SCs as possessed by the responders, and their character is a function from contexts in which the term's user is a responder, to the contents expressed by the term as used in such a context.

If KRAUT is opaque and applies to the individuals who cause the right sort of response in the right sort of subjects under the right sort of conditions, it can be deployed to refer to German people only in a context in which the speaker has certain negative (emotional, social) response to German people. KRAUT seems in this respect to parallel the behavior of the indexical concept I, that can be deployed to refer to an individual x only in a context in which x is the thinker herself.

We could in principle also imagine an indexical account where the speaker herself does not need to personally have the response (consider for instance accounts of taste predicates where they are indexical but need not include the speaker among the judges), but in this case it is not clear that the SC could still be opaque.

Because of the constraint on contexts carried by the terms' character, only responders (e. g. racists) can felicitously use the term, but non-racists can still understand the term by grasping its character. Understanding the term involves grasping its character; but only speakers in the right context (responders) can grasp the content of the term under its character. In other words, non-racists can't *think* the thoughts of racists who possess SCs, but they can *know* what thought they have. Deferential uses of STs by nonresponders might be blocked by the felt illegitimacy of the responders's attitudes towards the target - unlike the case of RED -, but still, nonresponders can *understand* the STs.

Possessing a concept is therefore sufficient but not necessary for mastering a term expressing that concept. Without even calling on deferential mechanism, one can be competent with a term without being able to deploy the relevant concept. This involves a distinction between two dimensions in concepts: character and content.

All competent speakers master the term because they have access to the concept's character (that is, they know that in the right context, the term/concept applies to those who triggered the appropriate non-conceptual state), but only the racists are in the appropriate context and can therefore access the concept's content.

Just like we can entertain demonstrative thoughts only in the right sort of (perceptual?) context, we can entertain pejorative thoughts only in the right sort of (emotional?) context. Yet, in any context, competent speakers can grasp the meaning (the character) of demonstrative or pejorative thoughts[112].

[112] Note that the analogy between SCs analyzed as reflexive response-dependent concepts (as indexical concepts) does not work all the way. In Perry's example of receiving a postcard saying "I am having a good time now", where the name of the writer and the date are erased (Perry 1993), intuitively, the addressee can grasp the character but not the content. The case of SCs is not intuitively like that. Even without the response, we can perfectly well say

Now that I have introduced the account, I will critically evaluate how RDAred handles the different explananda we started with, before raising some issues which will cast some doubt on the view and will eventually lead us to consider an alternative, non-reflexive response-dependent account of SCs.

which group is the (intended) referent, for we have independent, non-relational access to the (intended) referent that is the content of the expression.

Chapter 8. Pros and Cons of the RED Model

In the present chapter, I first evaluate how RDAred8 deals with most of the explananda we started with. We will see that it does quite well in accounting for major properties of SCs and STs such as hotness, expressivity, projection and so on.

Then, we will see the the analogy with secondary quality concepts might in the end have been drawn too quickly, as we will face serious disanalogies stemming from the nature of the response. The nature of the response will be the most problematic in trying to account for the possession conditions, and the categorization behavior of subjects.

This will lead us to qualify RDAred in reintroducing a conceptual element in the response, a conceptual element that will be responsible for categorization. We will consider the potential role of stereotypes in this respect, but eventually remark that this drastic move casts some doubt on the very project of RDAred to conceive of SCs as *reflexive* response-dependent concepts.

RDAred8 as an account of SCs as opaque and reflexive response-dependent concepts has important explanatory advantages:

- *Hotness*. First of all, RDAred provides a very satisfactory account of the intuitive hotness of SCs. Indeed, the emotional response yuk* is now constitutive of the concept (because we are in a reflexive case), rather than merely associated to it.

The response plays a central role: it is constitutive of the concept, in the sense that deployments of SCs are always closely associated to a response in the same way as deployments of RED are associated to perceptions of red. Such a close link to non-conceptual cognitive responses is a clear advantage of response-dependent accounts of SCs. The intuitive puzzle we began with was that SCs seemed to involve two kinds of content, a reference-fixing bit and a expressive or evaluative one.

Moreover, the two kinds of content seem to be somewhat inseparable, in virtue of reflexivity. With RDAred, we have an account of the close connection of the two kinds of content that goes further than all the hybrid accounts we investigated earlier. The two elements are more than merely associated: SCs are grounded on the emotional bit, which even helps identifying the targets.

That is so far the account of slurring representations which is taking hotness the most seriously. SCs are inherently hot, because they are grounded on a negative emotional response. No hot response, no (possession of the) concept.

- *Expressivity*. It is one thing that SCs are hot. But how come STs are expressive? An initially plausible hypothesis is that the expressivity of STs is the linguistic correlate of the hotness of SCs.

Given that SCs are hot in the way described above, and assuming my Conceptual Hypothesis (CH), it is likely that we will find traces of the inherent hotness of SCs in STs. Since SCs are grounded on negative emotional responses, uses of SCs (STs) consequently express concepts that are grounded on negative emotional responses. Expressivity here would be the expression of a concept loaded with emotions, or in short, the expression of emotions. Hence the feeling that STs are expressive. To see why such expressivity leaks out of most semantic operators, see the point just below.

- *Defectiveness/Projection*[113]*/Offense*. The offensiveness and projection of the expressivity of STs could be linked to the Defectiveness of SCs. Here is a plausible story that RDAred can tell about this important set of explananda. What is offensive in the use of (even embedded) STs could be that uses of STs show, through general pragmatic mechanisms, that the speaker is disposed to use such terms, hence that she possesses the concepts that these terms express.

But the simple fact of possessing concepts such as SCs is problematic, because possessing SCs requires, in virtue of reflexivity, having emotions such as contempt and hatred and so on (yuk*). That mechanism could be implemented by a simple non-specific conversational maxim like UP:

> **Use-Possession rule (UP)**: Language users usually possess the concepts that the terms they use express.

If UP is a general expectation in language use, that is, if even uses of "table" trigger the inference that speakers possess the concept of a TABLE, then uses of STs simply trigger an inference that speakers possess SCs. A version of UP specific to slurring representations would then be:

> **Use-Possession rule' (UP')**: Language users usually possess the SC that the ST they use express.

UP' seems like a good start to generate projection and offense. However, inference to the possession of an SC is arguably not sufficient[114]. Indeed, why would it be problematic and

[113] See the appendix to chapter 7 for a short study of the projection behavior of terms expressing response-dependent concepts under perspectival operators such as free indirect discourse.

[114] Nor is it specific to RDAred, incidentally. Most accounts subscribing to my Conceptual Hypothesis could explain projection and offense through an inference to the possession of the concept. That could be the case of potential presuppositional accounts of SCs (according to which SCs embbed a false presupposition about their targets), Hybrid expressivist accounts of SCs (according to which SCs are neutral concepts conventionally associated

offensive *per se* to merely *possess* a concept? I could possess a concept that is inherently bad (for short), but have *metacognitive* states about that concept that clear me.

For instance, I could know that the concept BOCHE I acknowledge possession of is a defectuous or "bad" concept, I could refrain from actively deploying it for this reason, and so on. I do not necessarily want to maintain that even this sort of possession is problematic and offensive.

A reply available to proponents of RDAred is that it is because of the special nature of SCs that mere possession becomes problematic. Indeed, for RDAred, possession presupposes a distinctive kind of response. So the possession of the concept, even supplemented with appropriate metacognitive attitudes, entails that the speaker has the response towards the target.

That is, for the speaker to possess BOCHE, she must have (or at least have had) a response of contempt/disgust/hate towards Germans. That is in itself problematic. We thus retrieve the inference to the speaker's negative attitudes from uses of STs. That requires an additional inference though (from the possession of the concept to the presence of a response).

So on top of inferring possession from use, hearers must be able to infer responsiveness from possession. That is, they must also follow an additional principle like PR:

Possession-Response rule for SCs (PR): Possessors of a SC are also responders.

This improved version of an account of projection is still problematic, though, in the three following ways.

A first observation is that it is still not clear why the response in itself is what is problematic. These are complex issues, but as a first approximation, we could argue that the relevant responses are typically at least partially automatic and involuntary. Being automatic and involuntary reactions, these cognitive responses would be out of reach of the subject's

with negative evaluations) and so on and so forth. As soon as the account is an account of concepts, and locate the source of the defectiveness and hotness in the concept itself, if it is admitted that slurring terms express slurring concepts, the option is open to account for projection in this way.

control: responses would be things that happen to them rather than things they do. And being out of reach of the subject's control, such responses cannot be the responsibility of subjects. SC possessors would be morally clean as they wouldn't have *done* anything wrong on that view.

There are at least two possible replies to this first observation. A first reply is that it is very unlikely that responses like "contempt" are completely out of reach of the agent's control. At least partially, contempt and hatred and negative evaluation of individuals *are* under the agent's control: we can introspect, dedicate efforts and discipline towards self improvement, work on one's bad habits and replace them with new ones, etc. So it is at least partly the responsibility of possessors (i. e. responders) that they have (kept alive) these sort of bigoted cognitive reactions to their fellow human beings.

Another and more simple reply to the observation that the response is automatic and involuntary is that even involuntary responses could be the source of moral/ethical/political/rational responsibility, or at least provoke offense.

My second point against the above account of projection stems from the fact that UP is qualified with a "usually", which wrongly predicts that offense should be cancellable. That is simply a general rule that speakers possess the concepts that the terms they use express, UP is more of a heuristic than an absolute rule. There are deferential uses of language where speakers do not have the mental representations associated with the terms they actually use, for instance.

With both UP and PR being mere heuristics, utterances of ST do not really *entail* having the response, and we could then cancel offense. In a context where it is sufficiently manifest that the utterer of an ST does not possess the concept, her utterance should not trigger projective derogatory content or offense. But as we saw, we do not find such non-derogatory uses are in the data[115].

Third, there is a potentially more problematic point in the account of projection we just considered: PR is in conflict with opacity, which goes directly against RDAred itself. Indeed, so as to draw the inference that the speaker is prejudiced from her use of an ST,

[115] To keep things simple, I leave clear quotational uses outside these considerations.

hearers must *i)* draw the inference from use to possession (UP) and *ii)* draw the inference from possession to response (PR).

But how can hearers infer that a speaker has a certain response from the fact that she possesses of a concept, without knowing that the concept is constituted by a response? If one infers the presence of a response from the possession of a concept in others, it must be because we possess some metasemantic knowledge of the RDB: we know that the concept is somehow tied to a response.

But Opacity claimed that these concepts, like RED, appeared subjectively just like primary quality concepts. From the subject's point of view, the concept applies to a property that is detected, and subjects are oblivious to the role that their own responses play in the construction of the concept and of its extension. There is here a conflict between PR and Opacity.

There is a potential answer to this last worry that proponents of RDAred could give, based on the distinction between two semantic projects. We distinguished earlier between two classes of relevant subjects in accounting for slurring representations: those who are the primary possessors and users (the *producers*) and those who know it only second hand, parasitically (some of the *consumers*). Racists who use the n-word among themselves are different from the rest of us who merely overhear their utterances and hence come to acquire some knowledge about the concept they express.

Hence, one thing is to uncover the semantics of SCs as they are possessed and deployed by the producers, another thing is to uncover the "semantics" of SCs as they are "possessed" by the other consumers. And with such a distinction at hand, RDAred could restrict its scope to the SCs of the producers.

In a linguistic community composed of producers only, there would be no such thing as projection nor offense, because SCs are opaque. Hence, speakers would simply deploy the concepts and use the terms to refer to their target, in a straightforward manner. Hearers would not make any particular inferences from the use of these terms, embedded or not. Since prejudiced against the target would already be common knowledge, uses of these terms would teach nothing to participants to the conversation.

Projection and offense arises only when two linguistic communities encounter. When a racist utters a ST in front of a non-racist, the non-racist learns something new about the user. The inferential mechanisms just sketched (relying on UP and PR) account for the sort of pragmatic enrichment that non-producers go through. SCs would be opaque for the primary deployers, but would have some degree of transparency for the other kinds of (parasitic) possessors. Correlatively, STs would be projective and offensive for the non-producers consumers, it would not be projective for the producers.

Such a reply might be fleshed out satisfactorily, but it is not necessary to do so, as there are alternative ways to trace back the source of projection and offensiveness. Projection and offensiveness could come not from the attribution of first-order states like possession of a concept or having a response, but rather from the implied absence of the right metacognitive attitudes towards SCs on the part of ST users.

As evoked just above, even if one could not help but possess SCs (one could have been unlucky enough to acquire language from a prejudiced community) one should strive to isolate the concept from the rest of one's mental life, refrain from deploying the concept, treat it as an inherently flawed concept, etc.

At the pragmatic level, this corresponds to an inference rule other than UP and PR. The defect of SCs would be not in the possession nor in the response, but in the not refraining from using SCs in thoughts, from the active deployment in ones conceptualization of the (social) world.

That inference could rely on a general maxim according to which we shall use only terms that express concepts we think are useful or adequate to actively deploy in ones mental life. We could call that plausible conversational principle the "Use-Approval rule" (UA). Here is a version specific to slurring representations:

> **Use-Approval rule (UA)**: Language users who are disposed to actually use STs are usually disposed to actively deploy in thought the SCs that the terms they use express, without reservations.

UA predicts that in cases in which use does not imply approval, offensiveness should be cancelled. This seems to be the case: arguably, a child who is manifestly parroting her peers, without understanding the undertones of the STs she uses is less likely to cause offense than the normal cases.

So it is really the possession of a term in ones active mental lexicon that triggers the projecting inference that the speaker is a bigot etc. This is why that content projects even under belief reports: indirect discourse relies on using our own language to describe something that was said by someone else (see Kaplan 1999). So the terms used must belong to ones dialect, and having a ST in ones dialect is a sign that one have the associated mental equipment[116].

- *Phenomenology*. Contempt is a well-studied negative moral emotion (see e. g. Bell 2013). Its phenomenology might be more complex than that of red*, but that is not necessarily problematic. The phenomenological aspect raises another issue that I discuss in the next section, though.

- *Dehumanization*. The dehumanization function of SC is according to RDAred derivative from the dehumanizing mode of thought triggered by the moral/emotional/social response.

Take the response to be one of contempt, to keep the argument simple. Arguably contempt itself dehumanize its targets. Feeling contempt towards an individual displays a lack of due respect, diminishes her value and importance as a person, and so on and so forth. Hence, it is not the having nor deploying of the SC that dehumanize in itself, but the contemptuous attitude that the SC is grounded on (Miller 1998).

- *Identifying thinking/Ideologies and Stereotypes*. RDAred does not come with any particular commitment regarding SCs identifying functions and their links to ideological and stereotypical thinking, but it is compatible with many kinds of independent accounts. It leaves open the possibility to let stereotypes play a significant role in enriching the response, so as to distinguish between the different responses involved in distinct SCs.

[116] See Hay 2011, who notices that "Jack believes that Pavarotti is a wop" expresses the negative attitudes of its speaker and does not attribute it to its subject (Jack), whereas "Jack believes that Pavarotti is a jerk" does the opposite. This could be because, under RDAred, having "jerk" in ones dialect is less indicative of psychological deviance than having "wop" in ones dialect.

- *Fineness of grain*. RDAred has the advantage of being very well armed to deal with the often overlooked phenomenon of coextensive slurring representations. For instance, NIGGER and SPADE are different SCs, but they share a common target. How can we account for their difference?

Most accounts have a hard time doing so, but RDAred could call on differences in the nature of the responses to the targets. There could be two groups of relevant subjects with two kinds of racist responses, and hence two SCs for the same target. As an extra advantage, RDAred has available a nice explanation of the puzzling fact that there are (almost) no cases of STs with a *positive* rather than a negative evaluative import (here are a few candidates: "aryan", "saint", the French "*savant*"): it could be a brute fact of ones social cognition that there is no natural kind of response to out-groups that is positive.

- *Contempt*. If SCs ought to be closely connected to emotions such as contempt, there is no closer link that the one RDAred provides.

- *Creation and evolution/Derogatory variation*. The existence and powers of STs is predicted to be as sensitive to change as the response involved in SCs. When society evolves and the group dynamics change, the emotional response of certain groups towards their targets might change. And since the power of STs arguably comes from such emotional responses, changes in the responses of the primary owners/users have direct consequences on the derogatory powers of STs.

- *Derogatory autonomy*. The inferential mechanisms accounting for offense and projection are automatic and conventional, so is the derogatory force of uses of STs.

- *Deference/Understanding*.

> We know what "Boche" means. We find racist and xenophobic abuse offensive because we understand it, not because we fail to do so. (Williamson 2009, worrying about Dummett's account in terms of rules of inference)

An additional advantage of RDAred lies in the fact that it comes equipped with a neat account of the shared understanding of STs across speakers, racists and non-racist. Both possessors and non-possessors of the SC understand the ST by grasping a *character* grounded on the biconditional[117].

Whoever understands "red" knows that it applies to objects (tomatoes, phonebooth etc.) that trigger a certain perceptual response on certain subjects. Such knowledge is visible in the mastery of the term.

The same holds of STs, whoever understands that the concepts it expresses applies to individuals that trigger a certain emotional/moral response in possessors of the concept. Crucially, this analysis of RDAred distinguishes between mastering a term and possessing a concept.

In *Reference, Inference and the Semantics of Pejoratives*, Williamson (Williamson 2009) criticizes Dummett for equating mastery of a slurring term with the disposition to use specific rules for the introduction and elimination of the term (Dummett, 1973). Williamson argues that non-racists understand racist terms even though they are not disposed to behave in accordance with the rules in question. He adds:

> Since understanding of the word "Boche" is presumably sufficient for having the concept that "Boche" expresses, it follows that a willingness or disposition to reason according to Dummett's rules is equally unnecessary for having that concept (Williamson 2009, p. 9)

RDAred agrees with Williamson that understanding a ST can be dissociated from the disposition to reason according to introduction and elimination rules, but should reject the premise that understanding a word is sufficient for having the associated concept. According to RDAred8, reflexive response-dependent concepts are context-dependent in the sense that subjects who are not in the right context can't deploy them.

Contrary to what Williamson suggests, however, understanding the word is one thing, possessing the concept is another. Just as you can't think of an object as "that thing" unless you stand in the right (e. g. perceptual) relation to the thing in question, you can't think of an

[117] Note that grasping a character does not equate possessing a concept, even though grasping a character provides with sorting abilities and so on. For more on concept possession, see chapter 9.

individual as a BOCHE unless you have the right kind of (emotional, attitudinal) relation to that individual or the class to which it belongs.

Still, every competent speaker understands the term "boche", even if - not being in the right context - they can't necessarily deploy the concept the term expresses: every competent speaker knows that, in a context where the emotional response is shared, the term denotes the individuals who provoke the response.

There is an important difference between color terms and STs, on this approach. In contrast to what happens with color terms, subjects who are not responders may not use a ST deferentially, because doing so would be endorsing the legitimacy of the responder's attitudes towards the targets. One does not defer to racists!

Still, non-racists can understand the term, as used by the responders (the racists). The non-racists know that the term refers to the objects that elicit the response in the racists (and they know what those objects are). The non-responders who master the linguistic term do not (and do not want to) possess the concept it expresses for the responders, but they know the biconditional that governs the deployment of the concept by the racists.

That means that the biconditional which governs the concept, although opaque for the possessors of the concept, is transparent to those who, not being responders, can't deploy the concept but still master the linguistic term expressing it.

These considerations pave the way for a distinction among cases of deference without possession. We will see in the next chapter that it would be odd to suppose that someone who has never tasted peanuts could not possess the concept PEANUT. This kind of case seems very different from a case of pure deference, like the layman's use of "elm", for which she has no (internal) sorting capacities and no associated concept.

But a proponent of RDAred8 could argue that making such a distinction would be in contradiction with the central idea that what grounds a concept is *constitutive* of the concept.

Such a reply would nonetheless be insufficient to draw distinctions such as the one between non-normal possession and non-possession *tout court*. What is it that distinguishes a subject who defers on judges from a subject who "defers" on her knowledge of the RDB in their application of a concept?

Both cases are non-normal, but one involves *possession* of a concept and the other does not. The difference might lie in the fact that non-normal possession provides sorting capacities grounded on systems *internal* to the subjects, whereas clear deference is completely external to the subject, which cannot count as a possessor.

8.2. Explanatory Cons of RDAred8

But despite its explanatory advantages, RDAred has non-negligible shortcomings. This will lead us to look for an improved version of a response-dependent account of slurring representations.

8.2.1. The problems of Extension and Phenomenology

- *Extension (the problem of non-emptiness)*. A consequence of RDAred is that SCs refer to the actual groups or individuals that actually trigger the yuk* response in NC-phobic subjects.

RDAred thus does not lead to an empty-extensionality thesis for SCs. That is, RDAred concludes that statements such as "There are Boches" or "Chomsky is a kike", and the thoughts they express, are true.

This could be problematic though, as it is hard to draw the line between ascribing the value True to "Chomsky is a kike" and actually holding the belief that Chomsky is a kike (see e. g. Richard 2008 or Hom and May 2013 for discussion).

A reply that proponents of RDAred could make is that what goes wrong in thoughts such as "Chomsky is a kike" is not the content but the vehicle. SCs are a flawed representation because they are grounded on an undue negative emotional or moral reaction to their targets, but still, they refer to their targets and can thus be true of them. Saying that it is true that "Chomsky is a kike" does not amount to saying "Chomsky is a kike", because the vehicles are irrelevant to truth and falsity. Content matters for metaphysics, not vehicles.

What the theorists wants to say when saying that "Chomsky is a kike" is true is simply that it is true that Chomsky is Jewish. The theorist, in her own (meta)language, would deploy JEWISH and not KIKE.

Proponents of RDAred should thus be careful not to subscribe to disquotational views on truth, but with that qualification, they might develop a non-racist view of SCs as non-empty concepts. This should not be problematic since disquotation does not apply to indexical statements, and STs are indexical under the proposed view.

- *Phenomenology*. An important difference between color concepts and SCs is that there does not seem to be a characteristic class of "racist" or "sexist" *qualia*, at least not on a par with color *qualia*. Inasmuch as it is natural to imagine that color concepts are normally grounded on one's color experience, it is less natural to picture slurring concepts as grounded on an alleged "racist" experience.

Appealing to an emotional response like contempt as opposed to lower-level, fully non-conceptual qualitative states somewhat addresses this issue. Yet this raises the issue of fineness of grain below.

8.2.2. The Problem of Circularity

Second, there an issue about a potential circularity of response-dependent biconditionals. When attempting to define a concept, one can't use the *analysandum* in the *analysans* (see Peacocke 1984, 1992 or Miller 2012).

A response-dependent biconditional is at risk of being trivial or circular when the biconditional is a necessary truth. Given a response-dependent biconditional, necessarily, everything that has the relevant dispositional property will trigger the right response in the right subjects under the right conditions, and anything that triggers the right response in the right subjects under the right conditions will henceforth have the dispositional property[118].

[118] There are non-reductive and reductive forms of a response-dependent bi-conditional. Lewis (1989) for instance proposes a response-dependent account of value, according to which something is a value if, and only if, we are disposed to desire having the desire for that thing under the right conditions (of imaginative acquaintance).

In the case of RED, when a response-dependent account argues that something counts as "red" if, and only if, it has the disposition to provoke a perception of red in the right kind of subjects under the right sort of conditions, the notion of "red" appears on both sides of the biconditional. An easy answer to the worry is that that the response red* can well be defined and characterized independent of the red property. It could for instance consist in a kind of neural response.

But it is less easy to offer a satisfactory answer to the worry of circularity in the case of SCs. The response yuk* can be characterized in terms of neural or physiological response of course, but the response must be sufficiently fine-grained to be able to distinguish between different SCs. And this can become problematic.

The relevant response for BOCHE must be different from the relevant response for KIKE, for BOCHE targets german people and KIKE Jewish people. yuk* must therefore be a germanophobic response in the case of BOCHE, and an anti-Semitic response in the case of KIKE. Everything looks like some conceptual element - the element distinguishing Jewish people from German people - must be brought back in the response.

Such a need to reintegrate concepts onto the primarily non-conceptual response would surely cast some doubt on the account. The extent to which RDA can stay truly response-dependent with a conceptual element in the response is an open question. Before being able to answer this question, we must first investigate further into the nature of the response.

Similarly, Johnston (1993) suggests that someone is responsible for an act if, and only if, we are disposed to hold that person responsible for that act. Both of these response-dependent accounts of the concepts of value and of responsibility could be seen as non-reductive response dependent accounts, in the sense that the extension determining response somehow involves or appeals to the concept being characterized. It is hard to characterize a notion such as that of "desire" independent on the notion of value, just as the notion of "holding someone responsible" is using the concept of responsibility itself. The extent to which these definitions are useful and interesting despite the degree of circularity they involve is debated. I shall from now on focus on reductive versions of response-dependent biconditionals, which do not raise these issues.

8.2.3. Conceptual Response

So a potential problem for RDAred has to do with the nature of the response yuk* - which should be further specified -, and more importantly should incorporate a conceptual element for it to be able to drive categorization and to distinguish between different SCs. Here are three reasons why.

1) Possession conditions

RDAred8 has the potential to tell a relatively detailed story about the possession of SCs. According to RDAred, the possessors of SCs would be the responders (in short the racists experiencing a yuk* response). Non-responders who could appear to possess SCs, just like the color-blind, would not count as possessors. So far so good.

Unfortunately for RDAred8, it seems that the posited class of possessors is too narrow, for there are many cases of deployments of SCs or uses of STs that would not count as such under RDAred8. Here are three cases of possession without response (that I call "cold possession") that I argue should be treated along other, maybe more standard, SCs and STs:

i) Picture Himmler as a "hyper-rational" Nazi, who may not have felt any particular emotions towards Jewish people, but was anti-Semitic to the extreme. His anti-Semitism could have been deprived of emotions, of any sort of yuk* response, and be fully confabulated: he believed that Jewish people were inferior and harmful beings, thought it was best for the common good to kill them all, and all sorts of outrageous and appalling anti-Semitic thoughts that Himler or other "intellectual" Nazis might have had.

It would be odd to suppose that Himmler lacked the SC simply because he lacked the yuk* response, to suppose that he had instead a neutral representation of Jewish individuals. He likely deployed KIKE when thinking about Jewish people, just like any other anti-Semite.

For RDAred to count Himler as a possessor of KIKE, the response must be more sophisticated than yuk*[119].

ii) Imagine a "benevolent" slave-owner in the 1800s who was not particularly emotional with regard to slaves, simply considered them as natural property. He called them "niggers" all day, and mentally represented them as inferior, less than human and so on. Why wouldn't the concept he deployed for his slaves be a SC?

Such a detached and careless attitude towards slaves might even have been the norm during slavery, and it is not at all given that this sort of cases asymmetrically depended on other cases of hateful and emotionally charged attitudes towards slaves.

It is conceivable that all possessors of the N-concept considered at that time their targets as objects and hence had no particular feelings towards them. But even in such a situation, their concept was a slurring representation, was harmful and flawed and dehumanizing and offensive in the same way as we know it to be today.

iii) So-called "country-club antisemites" seem to feel no particular animosity towards Jewish people[120]. They even ascribe primarily positive properties to them – they are smart, successful, etc. Again, even though all the attitudes they hold towards their targets are positive, it is still a case of anti-Semitism (for reasons that will become more clear later, having to do with essentialist thinking), and the concept they possess and deploy to target Jewish people should still count as a SC.

[119] RDAred8 could concede that Himler possessed the SC and notice that his concept of KIKE was likely to asymmetrically depend on the *hot* concept of the more primitive anti-Semites having yuk*. Or else, his own concept could have been acquired at a stage where he still had the emotional reaction to the targets, before suppressing his emotions and keeping the concept active by rationalizing it. These stories are sketchy but I wish to note that there are paths open to RDAred8.

[120] The example is due to R. May (pc.).

These three cases of "cold" (or emotionally positive) possession of SCs would entail that the simple "hot" negative emotional response yuk* should not be constitutive of SCs, contrary to what RDAred6 suggests[121].

The present counter-examples should thus be understood as pointing at the absence of real theoretical needs to keep "cold" and "hot" possession apart. Why should all cases of SC possession be necessarily hot? I do not see a reason other than an unmotivated will to give center stage to emotions in slurring representations. The response should thus be richer than what is suggested in RDAred8 with an emotional yuk*.

2) Fineness of grain

Take two color concepts such as RED and GREEN. Both are response-dependent and are thus governed by a RDB. The group of judges for both concepts are identical, it is the healthy well perceiving human beings. So what distinguishes GREEN from RED, that is, what makes the two concepts different concepts, must be the response. The response red* must be sufficiently distinct from the response green* for the concepts to target distinct properties of objects.

[121] We could in the three cases reply that these are precisely *not* possessors of SCs. On such a view, "cold" possessors would necessarily inherit degraded forms of SCs from "hot" possessors of original SCs. However, an argument for such a claim is required: at least *prima facie*, such cases do not necessarily involve deference, for instance. Whether or not these cases of "cold" possession really are cases of possession or not is partly a theoretical decision. It is not simple, and might be impossible, to find an independent set of criteria to pull apart cases of possession from other similar cases of non-possession or quasi-possession. There are cues such as the presence an asymmetric existential dependence relation between two kinds, but the split is also a matter of theoretical considerations such as simplicity.

Now, if we model SCs on color concepts, it follows that the negative emotional response to x - yuk* - is what cognitively distinguishes different SCs. The racist response must be fine-grained enough to be able to distinguish coextensional SCs (such as the N-concept and SPADE). But the response must also be fine-grained enough to distinguish between all SCs.

For instance, if a subject is both Germanophobic and Antisemitic, her Germanophobic response must be sufficiently distinct from her anti-Semitic response to cognitively distinguish her concept BOCHE from her concept KIKE. And that applies to all SCs. In other terms, in subjects who are diversely racists and prejudiced against different groups, there must be as many distinct cognitive responses of contempt and such as there are slurring concepts.

But it does not seem very plausible that there is a specific type of yuk* response for each specific SC, even less considering the fact that there is no well-identifiable phenomenology for racist experience. What looks more likely is that all kinds of racist responses are responses to "the other", and that what distinguishes SCs is their (independently characterized) target.

3) Categorization

But the most considerable problem for RDAred8 has to do with the apparent categorization behavior of SC possessors. In color perception and thought, RED-categorization is driven by red*-responses, naturally. That is, subjects under normal conditions detect that something is red in noticing that they see it red. They rely on their red* response to recognize that something is red and apply the concept RED to it. Red* responses drive RED-categorization.

By contrast, yuk*-responses are driven by GERMAN-categorization. Although the negative response could play a role, anti-Semites do not rely on their felt contempt to recognize Jewish people. Quite the opposite, it seems that they feel contempt because, and after, they (independently) recognized their targets as being Jewish. The relation between response and categorization seems reversed.

So for color concepts, the response comes logically and chronologically first and the categorization second, but for SCs, categorization likely comes independent of the response. This is a major difference between the cognitive role of RED and that of KIKE framed by RDAred8.

Because Germanophobes recognize their targets as being German independently of their negative emotional reactions to them, the yuk* response could involve the recognition of the target as a member of the target class. Here is a first refinement:

> **RDAred9**: x is a *SC* iff x would cause NC-phobic people to i) recognize x as a target, and ii) have a yuk* experience under [normal conditions of slurring thought].

So the relevant response playing the reference-fixing role in the RDB is now hybrid. It has *i)* an element of cold categorization, and *ii)* an element of hot response.

This first helps distinguishing between different SCs (addressing the problem of fineness of grain and possession conditions), because, for instance, although Germanophobes and Anti-Semites might have the same emotional yuk* experience of contempt/reject towards their respective targets, it is not the case that Germanophobes recognize Jewish people as their targets nor that Anti-Semites recognize German people as their targets. Such a more fine-grained and hybrid response will thus ascribe a different extension to BOCHE and to KIKE. The target of BOCHE is going to be the group of individuals that Germanophobes identify as targets (the germans) and reject/despise etc.

Additionally, having such a hybrid response helps account for the fact that subjects do not use their negative emotional response to recognize their targets (addressing the problem of categorization).

With the new version, subjects have on the one hand a recognitional mechanism to identify their targets, and on the other hand a negative emotional reaction to them. That is how subjects can have the first independent of the second. Whereas red objects are recognized by possessors of RED through their red* response, Germanophobic people recognize their targets (the Germans) through conceptual non-emotional routes, and have their negative response independently[122].

But at this stage we might wonder what role there is left to play for the negative emotional response, which was initially introduced to account for a real sort of hotness in SCs. If NC-phobic people can identify their targets through standard conceptual routes, couldn't we fix the extension of SCs as simply the set of individuals that NC-phobic people recognize as their targets? That is, couldn't the following version of RDAred work, without the second element in the hybrid response (ii) of RDAred work too? Consider:

> **RDAred10**: x is a *SC* iff x would cause NC-phobic people to recognize x as a target under [normal conditions of slurring thought].

It seems that such a version of RDAred concedes too much. Nothing would be left of the initial intuition we started with that SCs involved an inherent non-conceptual/emotional element. What is the point of having a response-dependent characterization of SCs if nothing is done to account for the intimate connection between SCs and emotional responses? Wasn't that a prime motivation?

Hopefully, to reinstate the non-cognitive response at its due central role, we need only to notice that the two components of the response in RDAred9 are not *independent*, after all. A two-layered response might in fact not be sufficient.

There must be a third element linking the first two, because NC-phobic subjects do not have the yuk* experience arbitrarily and independent of the recognition of their targets. It is on the opposite *because* they recognize their targets as German that germanophobes feel contempt for them. This is important for it is not the case that anyone who is German and is negatively evaluated by germanophobes will fall under the extension or target of the SC BOCHE.

For instance, the French germanophobe Jean could exclude Albert Einstein from the targets of his concept BOCHE (that would be a g-contracting deployment of the concept, as we saw earlier), but could feel a kind of contempt and reject for Albert Einstein on grounds others

[122] Recanati notes (pc.) that the two mechanisms could also interact, and that the emotional response could play a role in recognition and categorization.

than his membership to the group of Germans people. If Jean happened to be anti-Semitic on top of Germanophobic for instance, he would both recognize Einstein as a German and have the relevant kind of yuk* reaction to him, but still not apply BOCHE. To apply BOCHE, the response yuk* must be had *in virtue of* the recognition of the target as a German.

There should therefore be not two, but three components in the response: one responsible for the (possibly cold) recognition of the target, one responsible for the hot negative emotional reaction, and a link gluing the first two.

When a Germanophobe has a response to Germans, it is not that she takes the target to be simply, say, contemptible, nor that she takes the target to be German *and* contemptible, but rather that she takes the target to be contemptible *in virtue of* being German. The response involved in BOCHE is therefore not yuk* *simpliciter* like in RDAred8, not German-categorization *simpliciter* like in REDred10, not yuk* + German-categorization like in RDAred9, but something along the lines of RDAred11:

> **RDAred11**: x is a *SC* iff x would cause NC-phobic people to have a yuk*(i)-qua(ii)-NC(iii) response under [normal conditions of slurring thought].

The response is now threefold. It still has a negative emotional component yuk* (i), and adds a categorial component NC (iii) accounting for the recognitional capacities of subjects, and a link gluing the other two (ii). But what could be the link gluing the other two? What is it in Germans that germanophobes have a negative emotional response to? I have not talked much about the notion of a stereotype yet, but stereotypes are a likely candidate to play the role we need in the response. Germanophobes would take Germans to be contemptible in virtue of their having a certain number of (real or imaginary) properties.

I shall now make a short digression and consider the possibility that the additional conceptual element needed in the response is a stereotype.

So let us make a detour through the notion of stereotype and consider how it might be involved in SCs. I will mostly consider the view that SCs are stereotypes, but the discussion applies equally well to a view under which SCs are response-dependent concepts with a response involving a stereotype as the conceptual missing element noted above.

It was remarked that STs seem to go hand in hand with stereotypes (e. g. Jeshion 2013b or Saka 2007). Numerous theorists have indeed let stereotypes play some role or another in their account of the functioning of slurring terms. These include Dummett (1973) and Williamson (2009), who ascribe a function to certain stereotypes in inference rules governing slurs, but also Tirrell (1999), Camp (2011) or Croom (2011), in different varieties.

On the contrary some authors seem ready to put the stereotype in the semantic content of STs and SCs (hence in their sense). Hom (Hom 2008) for example suggests that the term "Chink" expresses a complex predicate of the form "ought to be subject to higher college admissions standards, and ought to be subject to exclusion from advancement to managerial positions, and … [insert other discriminatory practices], because of being slanty-eyed, and devious, and good-at-laundering, and … [insert other stereotypes], all because of being Chinese"[123].

Another example is Camp (2011), who insists that there are uses STs of where "the slur’s extension is restricted to stereotype conforming members".

[123] Incidentally, Hom argues that the content of this complex predicate is determined externally (like that of natural kind terms, which refer in virtue of the speaker's relations to the kind and to the linguistic community) and need not be represented by the possessors of the concepts or users of the terms, but that is another issue.

8.3.1. The Surface Stereotype View (SSV)

To see how stereotypes could play a role in shaping SCs, I start from an account that not only is intuitive, but also seems to provide a promising combination of simplicity and explanatory power – the *Surface Stereotype View* (SSV).

Intuitively, those who think of people as KIKES are applying a *negative stereotype* to Jewish people. Hence, the simplest account would appear to be that the concept KIKE *just is* a negative stereotype of Jews. More generally, SSV claims that slurring concepts are to be identified with negative stereotypes of their target categories.

But what are "negative stereotypes" exactly? While the phrase has entered common usage, the notion requires significant clarification in order to carry explanatory weight in a theoretical context.

There are different views of stereotypes. Putnam has a story on the stereotypes of natural kind concepts. He sets the notion of stereotype on its folk usage:

> In ordinary parlance, a "stereotype" is a conventional (frequently malicious) idea (which may be wildly inaccurate) of what X looks like or acts like or is. (Putnam 1975, p. 249)

On the one hand, for Putnam, knowledge of stereotypes is necessary to be said to have acquired the concept:

> Someone who knows what "tiger" means (or, as we have decided to say instead, has acquired the word “tiger”) is required to know that stereotypical tigers have stripes. (Putnam 1975, p. 250)

So we do not fully know what "gold" means unless we know that stereotypical gold is yellow. But on the other hand, importantly, this fact has no repercussion on the concept’s descriptive, reference-fixing component. In Putnam’s view, the reference-fixing role is external and deferred to the relevant experts.

There was no contradiction involved at all when chemists discovered that pure gold was in fact not yellow but white, even though stereotypical gold is yellow. Stereotypes are descriptive (e. g. "Gold is yellow" or "Tigers have stripes" are descriptions), but this descriptive material is not what plays the reference-fixing role for the concept.

Stereotypes, under this conception, are best seen as a "way of seeing" an independently fixed reference[124]. One conceives of tigers as being striped, even though unstriped tigers are conceivable. When we think of gold - or deploy the concept GOLD -, the property of yellowness is somehow involved in our thought without determining its reference, so that one can still conceive of non-yellow gold.

As opposed to Putnam's philosophical view on stereotypes, the most common view in psychology identifies "stereotypes" with sets of statistically weighted properties that members of the target group (are taken to) instantiate. For instance, flying seems to be more important of a property for birds than having two legs.

Under this conception of stereotypes, SCs could simply be usual concepts a set of stereotypical properties that the targeted class is taken to instantiate. Hom (2008, 2012) defends a version of such a view according to which slurs express complex predicates of the form "ought to be treated in *such and such* a way because of having *such and such* properties all because of belonging to *such* a group".

For instance, if it happened to be a statistically attested fact that German persons were particularly hard-working, offense could come from shortcuts in reasoning drawing conclusions about particular German individuals from such a generic premise. But having expectations towards an individual based on (one of the many) groups she belongs to is unfair (the individual could well be an outlier etc.).

Two questions relate to this view: what kind of features feature in the stereotype, and what role - if any - does the stereotype play in determining the extension of SCs? Stereotypes are not enough to define STs, as a set of e. g. positive stereotypes associated with a group do not seem to constitute a SC. The stereotype must be *negative*.

By "negative" stereotype, I mean that the features themselves are negative. Either all of the features are negative, or the set of features that the stereotype consist in are negative taken

[124] The notion of "way of seeing the reference" would better be worked out, but it is not indispensable to the following discussion.

together. But both ways, what it is for a certain feature or property to be negative needs to be spelled out more clearly. In line with RDAred, I will consider that for a feature to be negative consists in its typically provoking the relevant kind of negative emotional reaction in the relevant class of subjects.

8.3.2. Pros of the Stereotype View

SSV has several explanatory advantages. First, the existence of typicality effects in SCs could be seen as an argument in favor of their stereotypical structure of these concepts. The more stereotypical properties the target satisfies, the easier/faster subjects will be in categorization.

Plus, there seems to be gradable uses of STs, such as (171):

(171) !Einstein is more of a kike than Chomsky.

Under SSV, this would simply mean that one target satisfies a heavier set of stereotypical properties than the other[125].

Second, SSV deals well with the fact that SCs are *hot*. SCs are hot because the stereotypical properties happen to be negative, that is, to trigger some sort of a negative emotional or moral reaction in human subjects acquainted to them.

SSV can also offer an explanation of the *defectiveness* of SCs as a special case of the "representativity bias". There is indeed a potentially fruitful analogy to draw between stereotypical thought and "base rate neglect":

[125] But we shall not jump to conclusions, these uses might be due to something else. As soon as there are stereotypes associated with a concept, there are two different uses of the concept. Under one type of uses, the speaker/thinker focuses on the stereotypical properties and quantifies over them. Saying "you are a boche" will then be equated with "you have most stereotypical properties of Germans". But these uses are not Normal, in the important sense defined earlier. Normal uses are not gradable.

> Intuitive predictions are insensitive to the realiability of the evidence or to the prior probability of the outcome, in violation of the logic of statistical prediction. (…) people predict by representativeness” (Kahneman and Tversky 1973 p. 237).

Consider cruelty, a stereotypical property of Germans. It might well be that only 5% of Germans are cruel but that this rate is still higher than that of all other nationalities. But drawing on that basis an inference from German to cruelty is a faulty move, which corresponds to both stereotypical way of thinking and base rate neglect.

8.3.3. Cons of the Stereotype View

But SSV has explanatory gaps. First of all, if a stereotype is a set of features that members of the category have (or don't have) in different degrees, it must be said that it will be hard to stick to surface properties in the case of SCs. Indeed, under such an understanding of stereotypes as playing a role in determining extension, an individual who happens to have all the stereotypical features but as a matter of luck or by accident would be predicted to fall under the extension of the term. But this does not seem to be the case.

Take the concept KIKE for instance, and imagine a random set of stereotypical properties that anti-Semites assume characterize Jewish people (we let the reader figure out a random set of such terrible properties). It's easy to see how someone who is not Jewish could still satisfy all such properties that anti-Semites ascribe to Jewish people. But that does not make him or her Jewish, and hence he or she will not be thought of as a "kike" by racist thinkers.

So the SSV is not sufficient. And as the satisfaction of stereotypical properties does not seem to be sufficient to fall under the concept, it is likely that SCs are kind concepts. Note that this argument is a general argument one can raise against stereotypical views on concepts in general, it is not specific to the case of SCs.

> There are problems with equating concepts with undifferentiated clusters of properties and whit abandoning the idea that category membership may depend on intrinsically important, even if relatively inaccessible, features. (Medin and Ortony 1989, p. 179).

A stereotype view on the concept CAT would fall short similarly: a well-designed robot could have all the surface properties that humans detect to identify something as falling under the extension of CAT without thereby being a cat.

CAT applies to things that have a certain deep essential property (say the DNA of a cat), and as this property is inaccessible and invisible, humans use surface properties that they take to be causally connected to that essential property as reliable indicators that the perceived object does have the essential property, and hence falls under the extension. But this is, always, a fallible inference: essential properties are inherently invisible and hence always somehow postulated. I will come back to essences in chapter 10.

Now, it is clear that SCs are flawed, and we see how SSV accounts for that. But is it really sufficient to say that SCs are flawed because they inherit a representativity bias about the target? Is "boche" a bad word simply because it conveys the stereotype that german people are strict and cruel? It seems that we are missing an important component here, having to do with the dehumanizing power of SCs and STs.

Thinking of someone as a KIKE isn't simply statistically biased, it is reducing the persons' identity to a property, and fails to conceive of the target as a fellow human being. The person is seen through his or her category membership. This is why uttering even a neutral term like "woman", and even accompanied with positive stereotypes (e. g. "woman are better persons than men") is similarly prejudiced, causes similar reactions in the audience, is in the end a failed attempt at hiding one's misogyny.

If this observation is true, then SSV fails to fully account for the fact that SCs are flawed concepts. We need an extra component taking care of the dehumanizing power of SCs.

Finally, there are other reasons why a merely statistical view of stereotypes will not do. First, an individual could have all the stereotypical properties without falling under the extension of the concept. Conversely, some individuals might fall under the extension of the concept without having any of the stereotypical traits. That is, a statistical view of stereotypes will over-generate uses like "Of course you are lazy and so on, but you're not like the others, you're not a nigger", and under-generate uses such as "Obama is hard-working, sophisticated and civilized, but he is still a nigger"[126].

8.4 From RDAred to RDApolite

In sum, RDAred accounts for many important features of slurring representations, but it also faces serious and potentially insurmountable obstacles.

The main problem has to do with the need to reintroduce a conceptual element in the response so as to be able to explain the categorization behavior of possessors and the fine-grained differences between different SCs. Although it is conceivable to have a complex response of the form RDAred11 suggests (contemptible-qua-NC...), and to have a stereotype play the conceptual role needed to associate the negative response to the category, some doubt is now cast on RDAred.

Indeed, one of the main advantage of the view, based on the conjecture that SCs are opaque and reflexive response-dependent concepts, was that the non-conceptual, purely emotional, character of the response guaranteed as close a link between the concept and the response as was needed, and was able to derive the projection facts in an elegant manner (through inference to the possession of the concept, which was possible simply in virtue of reflexivity).

[126] Note to close this detour through stereotypes that Jeshion (Jeshion 2013b) forcefully argues against a stereotype semantics for STs. Here are some of her points:

"…much racism and bigotry is rooted simply in finding others "different" - often because of physical characteristics." (p. 322)

"There are bona fide slurs for groups for which there are not any corresponding societal stereotypes. Take the Yiddish "Goyim", used to refer pejoratively to all non-Jews, and 'Shiksa' to refer to non-Jewish women and girls. […] Without a societal stereotype to draw upon, the theory lacks the resources to explain the offensiveness of these terms." (p. 323)

But now that some conceptual element is back in the machinery, conceiving of SCs as reflexive is less appealing. This is why I will now consider a view of response-dependent concepts as non-reflexive, and put forward another response-dependent model of SCs based on concepts such as POLITE and COMFORTABLE.

Chapter 9. A Non-reflexive Response-Dependent Account

In the present chapter, I reexamine the view that response-dependent concepts such as RED are reflexive concepts. Distinguishing between two notions of concepts, one fine-grained and one coarse-grained, I argue that concepts shall not be individuated by the cognitive means that are put in place to ground them.

Under this view, color-blind people can, in principle, possess the "same" concept of RED as the non-color blind. The way in which different possessors differ in their grounding of the concept (e. g. RED is grounded on red* in well-seeing subjects, and may be grounded on knowledge of the RDB in color-blind subjects) shall not play a role in concept individuation, I shall argue.

I will then investigate another response-dependent account of slurring concepts based on this updated conception of response-dependence, on the model of the concept POLITE. We will see that this last move has dramatic consequences, such as the disappearance of the clear-cut distinction between slurring concepts and their neutral counterparts, not to say the dismantling of our central notion of "slurring concepts" altogether.

Let me now endeavor to show that response-dependent concepts, contrary to what seemed to have been established earlier, are *not* reflexive; even color concepts like RED. I want to argue that the very notion of Reflexivity is committed to a vision on the individuation of concepts that is not appropriate. Concepts shall not be individuated on the cognitive mechanisms that are effectively used to apply them adequately, I shall argue. It must therefore be *possible* to possess any concept, be it a response-dependent concept, without having the appropriate cognitive response.

Consider first the view that RED is Reflexive:

> **Reflexivity Thesis** (for RED): it is not possible to possess RED without having the response red*.

Under that version of the Reflexivity Thesis, red* is constitutive of the concept RED. For a certain concept to count as the concept RED, it is necessary that it involves the perception of red in the right manner. A certain subject who would display capacities to sort red objects from non-red objects similar to that of normal possessors of RED would thus not be counted as possessing the concept RED.

What then can account for these cognitive abilities that are behaviorally indistinguishable from the everyday use of "red"[127]? Proponents of the reflexivity thesis would argue that the omniscient color-blind who perfectly ascribes "red" to red objects does not possess RED but another concept. Call this concept RED'.

RED', as relying on knowledge of the RED RDB, is a descriptivisation of RED constructed out of the RDB. RED and RED' are two different modes of presentation of the same property, one applies in virtue of its causal connection to cognitive responses, the other

[127] Note that a proponent of fine-grained concepts would not concede that these difference abilities are behaviorally indistinguishable. It is even precisely because of this that they recommend individuating concepts in a fine-grained manner: because behaviors are different even if concepts are coextensional.

applies in virtue of its encoded semantic content, roughly of the form "the property that is responsible for the cognitive response red* felt by healthy human beings in normal lighting conditions". Summing up, there are thus two coextensive concepts that apply to red objects:

> **RED**: The intuitive concept, grounded on the response red*.
>
> **RED'**: A descriptive concept, grounded on knowledge of the RDB.

The reflexivity thesis thus accounts for the application of "red" by omniscient color-blinds by postulating the existence of another mental representation of red, another concept, RED'. We could instead argue not that there are two different concepts, RED and RED', but that there are two *ways* to ground ones RED concept.

The same would apply to SCs. To explain the differences between SCs as possessed by the racists and their correlates as understood by others, RDAred could try to make a distinction between concepts possessed in a *Normal* way, that is grounded on the response yuk*, and concepts *non-Normally* grounded on knowledge of a RDB for instance.

By "normal possession", RDAred could rely on our earlier discussion of Millikan's notions, and mean the following:

> **Normal possession**: The subject's possession of a concept is *Normal* when it relies on the conditions under which the fulfillment of its proper function allowed for its reproductive success[128].

The view would account for any other kind of non-normal applications of the concept in the same vein: whenever a subject deprived of the response uses "red" or seems to mentally represent redness, she does so in virtue of her mastery and possession of another, possibly coextensive, concept. That concept, not being grounded on the subject's response red*, is not identical to RED, because red* is constitutive of RED.

On the model of RED, we can consider a more general version of the Reflexivity Thesis, which would apply to all empirical concepts, that is to all concepts that are grounded in the same way as RED on a certain (cluster of) cognitive response:

[128] The color-blind color-scientist's possession of the concept RED is thus not Normal. The Normal possession of RED is grounded on red*, not on scientific expertise.

> **Generalized Reflexivity Thesis**: possession of an empirical concept F requires having the associated canonical response R.

It is conceivable that the concept CROISSANT is an empirical concept, in the sense that there might be a canonical mode of presentation of croissants. This might be a combination of visual, gustative and olfactive impressions for instance. Here is a tentative definition of canonical modes of presentation, a correlate notion of Normal Possession:

> **Canonical Mode of Presentation**: a concept's canonical mode of presentation is the Normal mechanism it is grounded on and which provides its Normal possession conditions.

Maybe that this canonical way of being introduced to croissants is constitutive of the concept CROISSANT, in the sense that someone who has never seen or felt or tasted a croissant could not really possess the same concept as mine or that of other croissants enthusiasts.

Under the Reflexivity Thesis, then, such a subject, even if competent with the term "croissant", does not possess CROISSANT but a different coextensive concept CROISSANT' (see Williamson 2003).

9.2. Non-Reflexive Concept Possession

I will now argue in favor of a Non-Reflexive understanding of response-dependent concepts in general, before introducing a non-reflexive response-dependent view of SCs - RDApolite - and evaluate its advantages and shortcomings.

9.2.1. Recognitional Concepts

The Reflexivity Thesis presupposes what Fodor calls "Empiricism", that is, the view according to which

> the content of at least some concepts is constituted, at least in part, by their connections to percepts. (Fodor 1998, p. 2)

In other words, the Generalized Reflexive view that all concepts are in some sense connected to perceptual or other cognitive responses rests on the assumption that concepts can be individuated by epistemic properties.

Fodor argues that there can be no such concepts (concepts he calls "recognitional concepts" in Fodor 1998a), and that concepts are solely individuated by the cognitive ability of agents to distinguish things that fall under the extension of the concept from things that do not (see Recanati 2002a for a critique of Fodor's argument against the existence of recognitional concepts).

Why should we suppose that there are two concepts of red (RED and RED', for the intuitive case and the colorblind case)? Why should we suppose that colorblind people have a different concept of RED than others?

One argument in favor of that view relies on the asymmetric existential dependency there is between the color-blinds' concept and the non-color-blinds' concept. In a world without any colorblind people, subjects would still possess and deploy the concept RED, whereas in a world of colorblind people, there would be no use for the concept RED. The colorblind's concept is thus parasitic on the non-colorblind's concept, which could be taken as a clue that the two concepts are structurally different.

But a clue is not enough here. RED and RED' might be different versions of one and the same concept, even though one is parasitic on the other. For instance, my concept of an IPHONE depends existentially, asymmetrically, on that of Steve Jobs', but that does not mean that it is a different concept of IPHONE. Otherwise only Steve Jobs would have ever had the only true concept of an IPHONE, and everybody else would possess some different concept. That is not plausible. It seems as if pursuing the argument further would lead us to a view of concepts that is so fine-grained that all of the concepts we acquired through communication would be unique and distinct from any other.

Another argument in favor of the Reflexivity Thesis relies on Frege's criterion for concept individuation. There must be two distinct concepts of red, one intuitive (RED) and one theoretical (RED'), because we can construct a Frege case where "Tomatoes are red" is assertable and "Tomatoes are red'" is not.

If one is a healthy naive subject for instance, we might assert that tomatoes are red without thereby being ready to assert that tomatoes have the right micro-structural properties to cause normal human perceptual systems to experience a sensation of redness under normal lighting conditions.

But according to Frege's, the substituted elements must have distinct semantic content. Hence there must be two distinct concepts of red, in accordance with the Reflexivity Thesis.

But the Fregean argument for the Reflexivity Thesis is objectionable. Similar to the first argument based on asymmetrical dependency, it seems that it would lead us to have too fine-grained a view on concept individuation, even more fine-grained than de fineness of grain distinguishing RED from RED'. There would in fact be infinitely many concepts of RED, that is, as many concepts of RED as there are conceivable Frege cases involving "red". There would be as many concepts as there are Frege cases, and this might be too much. The criterion for concept individuation suggested by reflexivity thus seems to be too fin-grained.

But any sort of a cognitive difference related to RED can lead to a Frege case. Imagine that Frank believes that all red objects come from Mars. There is thus a cognitive difference between Frank's concept of RED (call it RED1) and the more usual concept of RED which does not embed that belief.

Imagine you present Frank's with a counterexample to his belief, for instance a red stone that you can prove comes from the earth. In that situation, Frank knows that the red stone he faces does not come from mars, and thus admits that it is false that "the stone is red1". Frank would thus be led to form a novel concept of red, RED2, that does not embed his prior false belief[129].

It is in fact a common criticism to Frege's criterium that it leads to an arbitrary level of fineness of gain for concept individuation, as even synonyms such as BACHELOR and UNMARRIED MEN can give rise to Frege cases (see e. g. Mates 1969, or Sainsbury & Tye 2013).

But surely anything that distinguishes synonyms must be more fine-grained than meaning[130]. So Frege cases should not be taken at face value in weighting the Reflexivity Thesis either, I argue. Again, the goal is to construct a response-dependent account of SCs that would

[129] If the paradigmatic case of RED is not reflexive, what response-dependent concept could be? A potential example of a reflexive RD concept, which I owe to Benjamin Spector, could be the French MIGNON. MIGNON applies to male and roughly means "attractive".

It seems that utterances of "mignon" systematically trigger the inference that the utterer is attracted by men (in non-deferential cases), as if the concept MIGNON was inherently or constitutively linked to a certain kind of response to men. SCs could then be the MIGNON sort of concept rather than the RED sort of concept then.

But the reasons to reject the idea that cognitive responses individuate concepts are more general though. Patterns of inferences to the effect that the speaker is attracted by men could follow from the fact that being attracted to men is the Normal possession condition for MIGNON. That does not close the door to potential non-responder non-deferential possessors of MIGNON.

[130] F. Recanati suggested to me that there are two other possible conclusions of Mates' argument (p.c.). One is that there are no real synonyms. Another is that cognitive significance is not a purely semantic matter, but also involves the vehicle.

account for the specificity of the content of SC. An account that would not succeed in doing so is insufficient to characterize SCs.

At the end of the two arguments in favor of the Reflexivity Thesis, we see that categorization behavior and concept individuation are two different things that it is better keeping distinct. The same dispositions to categorize objects appear to be neither necessary nor sufficient for possession of the same concept.

It seems that the only theoretical question there is left is whether or not we need super-fine-grained concepts? I do not see why we would, if not to account for differences in categorization behavior.

There might be differences, even dramatic differences, in the way the two subjects possess and master F, in the sort of environmental relation to the reference they ground F on, in the kind of inferences it licenses for them, or the sort of non-conceptual cognitive activity it is linked to. But not all such cognitive and environmental differences do necessarily count as differences between the concepts (see Williamson 2003).

Concept could be more simple, and based on the subjects abilities to have thoughts about the concept's referent, in one way or another, as Pettit emphasizes:

> A person has a concept of something, I hold, if and only if she is able to try to form rational and true beliefs that bear on that thing. (Pettit 1991, p. 595)

The following example illustrates that intuition. Imagine that the taste of peanuts is a canonical mode of presentation of peanuts, so as to be constitutive of the concept PEANUT. The generalized Reflexivity Thesis about PEANUT holds that someone who has never tasted a peanut does not possess a PEANUT concept.

Now, many people are allergic to peanuts and some of have never been able to taste a peanut. But still, they say things like "I am allergic to peanuts", "If I eat a peanut it will burn my tongue and throat", "the taste of peanut would not be the same for me as for you" and so on and so forth. They think about peanuts, try to avoid being in contact with peanuts, talk about peanuts. There is no real motivation to say that they do not possess PEANUT, other that the mere will to ground concepts on canonical modes of presentations[131].

And an account that counts PEANUT as response-dependent fails to respect the requirement that SCs be characteristically response-dependent. We do not want to know that SCs are response-dependent in the way any other empirical concept is, we want to know in which specifically interesting manner SCs are response-dependent. Reflexivity does not appear to give the wanted specificity.

An answer the Reflexive Theorist could give in favor of more fined grained concepts is that coarse grained concepts can be formed after fined-grained ones, so that having extremely fine-grained concepts is not theoretically problematic. I grant that point. Postulating extremely fined grained concepts is not theoretically inadequate *per se*. My point is rather that such a postulate lacks a clear motivation other than Frege's criterion. As long as it is left unmotivated, I will prefer neglecting it altogether.

The Reflexivity Thesis should thus be supported by an independent argument in favor of a (theoretically priviledged) type of concepts whose possession requires a response. This seems crucial to the Reflexivity Thesis. And the independent argument cannot simply be: "Intuitively, hotness is crucial", as there are many other ways to account for hotness.

A last limit to the Reflexivity Thesis is that it does not seem well-equipped to distinguish between cases of possessors who master the descriptive concept via their knowledge of the RDB, and deferential deployers who do not possess the concept at all. Indeed, subjects who are not responders may still use the term expressing a color concept such as RED, even if, qua non-responders, they don't possess the concept. They can do so in virtue of the mechanism of deference at work in language use (Putnam, 1975):

> Everyone to whom gold is important for any reason has to acquire the word "gold"; but he does not have to acquire the method of recognizing if something is or is not gold. He can rely on a special sub-class of speakers. The features that are generally thought to be present in connection with a general name—necessary and sufficient conditions for membership in the extension, ways of recognizing if something is in

[131] One could argue that that there is not one, but many different PEANUT concepts.

> the extension ("criteria"), etc.—are all present in the linguistic community considered as a collective body; but that collective body divides the "labor" of knowing and employing these various parts of the 'meaning' of "gold". (Putnam, 1975, pp. 227-8)

For deferential users, the word "red" refers, for them, to the property which possessors of the concept (i. e. responders) refer to when they deploy the concept, namely the property of falling into the class of objects that elicit the response in them. Thus, when saying that "tomatoes are red", colorblind people say that tomatoes are whatever it is that the judges for color see red. That is hardly distinguishable from the possession of RED' the descriptivisation of RED. There needs to be a distinction between mere deferential possession and cases of possession by non-responders[132].

9.2.2. Non-Reflexivity

So suppose we make instead few demands on concept possession, and that e. g. a color-blind person can be said to possess the concept RED simply by being appropriately situated in a speech community of color-sighted people. This view entails to the Non-Reflexivity Thesis:

> **Non-Reflexivity Thesis (for RED)**: the response red* is not a necessary condition for the possession of RED, it is simply its Canonical Mode of Presentation.

Under that view, the competent color-blind color-scientist possesses RED on a par with healthy subjects. To extend that view of concept possession (hence individuation) to other empirical concepts such as PEANUT, consider:

> **Generalized Non-Reflexivity Thesis**: No canonical mode of presentation of F is a necessary condition on the possession of F.

[132] F. Recanati remarks that there are two senses in which we talk of deference, one in which categorization is possible and one in which it is not (p.c.). Discussions of deferential concepts often overlook this difference. For more on this debate, see Recanati's response to Woddfield in Recanati (2000).

The Generalized Non-Reflexivity Thesis thus posits a notion of concepts that is somewhat coarse grained, as it refuses to individuate them on the basis of cognitive responses. It does not undermine the role of cognitive responses though, as it recognizes that concepts usually have a canonical mode of presentation. The canonical mode of presentation of a concept plays a non-negligible role, but does not impose possession conditions[133].

Under that view, we can address the last worry presented above, and distinguish between cases which rely on deference and Normal cases. Normal subjects and color-blind scientists who master the RED RDB both possess the concept RED, and uses of "red" by color-blind subjects who do not master the RDB are just deferential cases without concept possession.

Note that when it is combined with a view of concepts as psychological entities individuated by the cognitive sorting abilities of subjects, the notion of deference without concept possession becomes problematic.

Indeed, if the only criterion to decide whether a certain subject S possesses the concept F or no is his ability to sort Fs from non-Fs, it should be acknowledged that even deferential users are able to sort: they can ask judges. Under some sense of "ability", they have the right sorting ability. Colorblind do know that tomatoes are red and bananas are not red, and faced with any object, they can ask a member of the group of color judges to help them. That is the sense of Fodor's remark[134]:

> I can't tell elms from beeches, so I defer to the experts. Compare: "I can't tell acids from bases, so I defer to the litmus paper"; or "I can't tell Tuesdays from Wednesdays, so I defer to the calendar." These three ways of putting the case are, I

[133] F. Recanati notes that a bad consequence of this theory is that there are no demonstrative concepts (p.c.).

[134] See also Laurence and Margolis's:

> As long as tokens of proton are suitably connected to protons, it doesn't matter how the connection is sustained [...]. You can have your beliefs and I can have mine, and the differences in our beliefs won't in themselves entail that we are subject to conceptual differences. Whether our concepts are different depends upon their connections to the world. (Laurence and Margolis, 1998, p. 353)

> think, equally loopy, and for much the same reason. As a matter of fact, I can tell acids from bases; I use the litmus test to do so. And I can tell elms from beeches too. The way I do it is, I consult a botanist. (Fodor, 1994, pp. 34-5)

So even though the Non-Reflexivity Thesis does better in i) avoiding unnecessarily fine-grained concepts and ii) accounting for the difference between deferential cases and Normal cases, it should be supplemented with a more detailed account of deference without concept possession.

A possible answer to the problem of deference without possession could be that concept possession requires not simply sorting abilities, but sorting abilities that are grounded on an *internal*, rather than external sustaining mechanism. When a subject has the sorting abilities she has in virtue of certain relations of her concept with her own cognitive system, she could be said to possess the concept. When the link connecting the concept with the objects it refers to is external to the subject, though, it would count as a deferential case. The extent to which such a view concedes to a fine-grained understanding of concepts could then be further investigated.

Apply this notion to RED again. Of course, if I happen to be sensitive in similar ways to a ripe tomato and a phone booth, I can rely on the similarity of these contingent perceptual responses to form a concept applying to both a ripe tomato and a London phone booth.

But that is not to say that perceptual sensitivity is the only logically possible way to acquire and master the concept RED; a person can possess the same color concepts as others, without relying on the response, even though it might happen rarely because of the cognitive cost of doing so.

The usual normal way to acquire and possess the concept RED is by relying on the perceptual system, but it need not be the only way. My insensitivity to ultrasound or infrared does not prevent me from having a concept for these things, but I will have to rely on something other that my own perceptual responses to be able to apply these concepts correctly: for example an external measuring device or an expert's testimony. What I rely on to connect a concept with its referents is a sustaining mechanism:

> .. acquiring a concept involves establishing a sustaining mechanism that connects the concept with the property it expresses. (Laurence and Margolis, 1998, p. 359)

Figure 8 represents the "Normal" sustaining mechanism for RED, that is, the one where the concept is grounded on the response red*:

FIGURE 8. A "NORMAL" SUSTAINING MECHANISM FOR RED

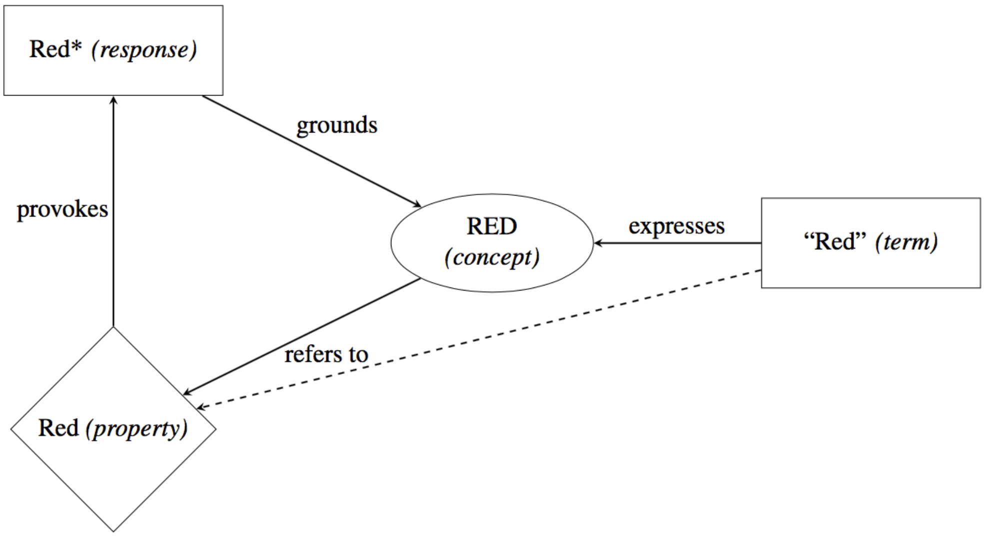
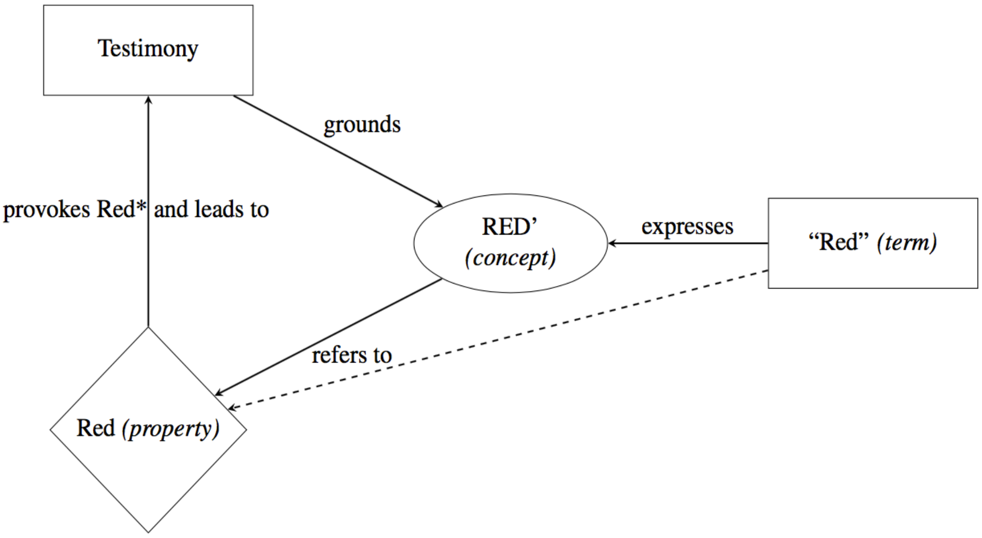


Figure 9 represents a deferential sustaining mechanism for RED. As a Non-Reflexive response-dependent concept, it is still the same concept RED, but this time grounded on the testimony of the color judges:

FIGURE 9. A DEFERENTIAL SUSTAINING MECHANISM FOR RED

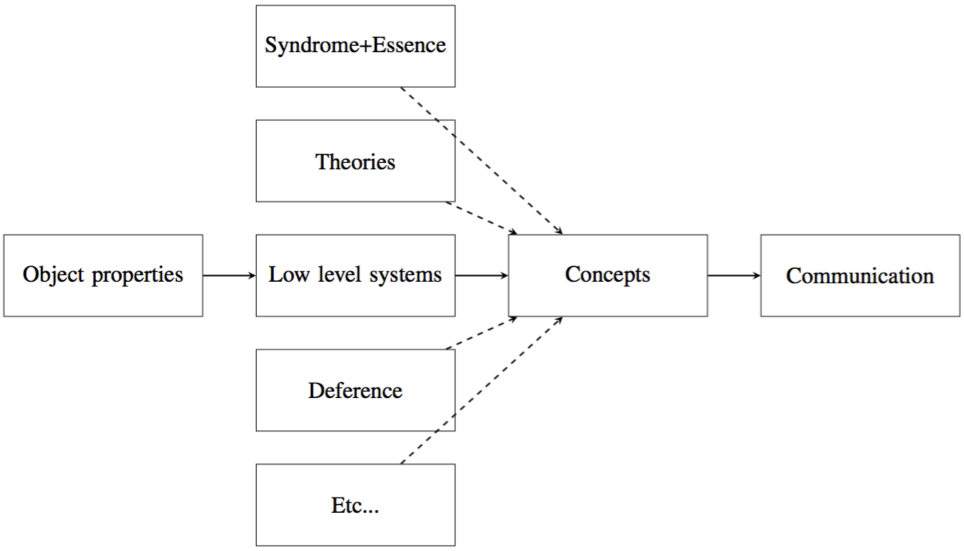


In the end, the conception of concepts we arrive at is the following. There exists canonical modes of presentation of most empirical concepts, and these canonical modes of presentation usually involve the responses of the subject's cognitive system. In the case of color concepts, it will be responses of the perceptual system.

But low-level systems are not the only systems feeding concepts. Even if cognitive responses are possibly the canonical sustaining mechanism, there are many others. We saw deferential sustaining mechanisms, but there are others: a concept can be a purely theoretical concept (maybe SQUARE) and be applied independent of cognitive responses, or it can be an essentialist concept (maybe CAT) in the sense that it is applied when subjects are led to postulate an essence in the object based on the detection of a syndrome[135], and so on and so forth.

Figure 10 represents how concepts are created by ones cognitive system after the detection of object properties, and hence feed communication:

FIGURE 10. FROM PROPERTIES TO COMMUNICATION



Recall that a potentially negative consequence of non-reflexivity is the identity between RED and RED' above. This is a bad consequence because, applied to SCs, it identifies SCs as they are deployed by racists and the deployments of SCs by non-responders. But if it is possible to possess SCs without being a responder, it is unclear to what extent SCs exist as a natural kind of concepts. Any one could, in principle, non-problematically possess them.

[135] More on essences in chapter 10.

9.3. Slurring Concepts Modeled on POLITE

Even with the problems encountered earlier, there is still some room for a response-dependent account of SCs, because not all response-dependent concepts are similar to RED. We can locate the source of the most problematic issues that RDAred faces in its reflexivity constraint.

9.3.1. Introducing RDApolite

Indeed, only if SCs are reflexive response-dependent concepts must their possessors rely on their negative emotions for categorization, must themselves be "hot" possessors for instance. Were opacity put aside, there would be room for cold Normal possession - for racist possessors grounding their SC on their knowledge of the RDB - and for categorization independently of a response.

So maybe that the model of color we started with - on the intuition that it was the best way to take hotness and expressivity seriously - was not the right response-dependent model. Let us consider a model of SCs based on a non-reflexive response-dependent concept instead, such as POLITE[136].

POLITE is likely to be a response-dependent concept, because it is clear that its extension crucially depends on the responses of certain subjects. It would be very difficult to give a set of necessary and sufficient conditions for "polite" behavior.

What is common between welcoming someone, keeping one's hands visible when eating, or holding doors? It makes no sense to gather these completely different kinds of events independent of the fact that certain human beings tend to respond to them in a common manner. So polite things are polite in virtue of being disposed to cause a certain kind of

[136] Many thanks to Aidan Gray for the suggestion (p.c.).

polite* response, in a certain group of subjects, under the right kind of circumstances. POLITE is thus governed by a RDB such as the following:

> **POLITE**: x is *polite* iff x would cause educated human beings of the community to have a polite* judgment under normal social conditions.

I describe the relevant group of judges for POLITE as "educated human beings of the community" because knowing what is polite and what is impolite arguably requires some form of (explicit and/or implicit) training, as politeness is usually conventional. And precisely because politeness has a conventional dimension, what counts as polite in a country could count as impolite in another[137].

The set of judges should therefore be relativized to communities. I call the relevant response a "polite* judgment". "Polite*" so as to avoid circularity, again, and "judgment" because seeing something as polite intuitively looks more conceptual than a pure cognitive response such as color perception. That is debatable of course, but nothing crucial will hinge on these assumptions.

Finally, I chose "Normal social conditions" as a placeholder for the relevant conditions. Intuitively, the conditions under which polite judgments hold are everyday conditions where you can expect others to be civilized and so on. We would discard war-like situations in judging polite behaviors for instance, or panicking crowds and other such extreme or less extreme situations.

Concepts like POLITE are importantly different from secondary quality concepts in (at least) the three following ways:

[137] This entails that someone who knows the conventions could be able to state the low-level properties which make a certain piece of behavior a "polite" behavior, which corresponds in the end to a descriptive - not response-dependent - concept of POLITE.

1) Control and coordination

Unlike red*, the extension-determining response polite* functions to coordinate behavior in groups. Red* is a low-level perceptual response that is very stable cross-culturally, automatic, and out of reach of the subject's control.

Subjects thus coordinate on the terms expressing the color concepts they built out of their responses, but they cannot modulate the response itself. But it seems that being more conceptual and high-level, the response polite* is less automatic and more accessible to the individual subject's control.

As a consequence, subjects can become more or less responsive to certain kinds of behavior and coordinate on the kind of behavior they collectively want to be responsive to. That is why there is more cross-cultural variation in politeness-judgments for instance, and why what counts as polite is in a sense more arbitrary than what counts as green or red. In short, the extension of POLITE is at least partially conventionally decided by the community, where the extension of RED imposes itself to the subject and community.

Just like COMFORTABLE, POLITE is not an opaque response-dependent concept but a transparent one. This has two other consequences: one about the *a prioricity* of the RDB, another about conditions of (Normal) possession.

2) A prioricity

Subjects who (Normally) possess the concept POLITE must know that it applies to pieces of behaviors that are commonly judged to be polite. They do not take politeness as a primary quality that they detect in pieces of behavior.

When I judge that not answering to an email is impolite, I judge so because I know that not answering emails provokes the kind of negative reactions and judgements associated with impoliteness, that is, I have an understanding of the intimate connection there is between (im)polite behavior and people's reactions to such behavior.

That was not the case of RED though, that I could master and apply without awareness of the intimate connection between redness and people's perception of redness. Both RED and POLITE are governed by a RDB, but knowledge of the RDB is not Normal in the case of RED, whereas it is Normal in the case of POLITE. Knowledge of the RDB is part and parcel of (Normal) possession of the concept, it is thus (Normally) *a priori*.

3) Non-responder possession

The transparency of POLITE also has a consequence on the requirements on Normal concept possession. Because knowledge of the RDB suffices for Normal possession, non-responders can normally possess the concept and token-apply it. For instance, I could be totally insensitive to someone's acknowledgments, and still recognize her words as a polite piece of behavior.

A first simple application of the POLITE model to SCs would give a RDB similar to RDAred, but with the important difference that it would be a transparent rather than opaque RDB:

> **RDApolite***:* it is (Normally) *a priori* that x is a SC iff x would cause NC-phobic people to have a yuk* experience under Normal conditions of slurring thought.

Let us now consider how RDApolite deals with the criticisms to RDAred.

- *Extension (the problem of non-emptiness)*. RDApolite as it stands still predicts the non-emptiness of SC. Indeed, it is the case that certain individuals trigger a "contemptible-qua-german response in Germanophobic people under the appropriate kind of conditions. These people are going to fall under the extension of BOCHE, then, and it will therefore be true that there are boches.

One possible way to derive the favored emptiness in an RDApolite theory of SCs could be to reformulate the "yuk*" bit of the response in modal terms such as "contemptible" or moral terms such as "worthy of contempt". Indeed, although some individuals do actually trigger a contempt-*qua*-German (or yuk*-*qua*-German) response in Germanophobic people, nobody is actually worthy of contempt because of belonging to a group. And since nobody is "contemptible" or "worthy of contempt", SCs end up with an empty extension.

Let us thus consider RDApolite, the final version of the view (we keep yuk* for the emotional bit, as it is still unclear whether contempt is sufficient):

> **RDApolite**: it is *a priori* that x is a SC iff x would cause NC-phobic people to have a worthy-of-yuk*(i)-*qua*(ii)-NC(iii) response under Normal conditions of slurring thought.

This version of RDApolite is thus well armed to address the issues it was designed to address while ascribing an empty extension to SCs, provided that deriving emptyness is a requirement for a theory of SCs.

We could surely apply the same move to RDAred and move back to a reflexive response-dependent model, but we would still face the important problem of categorization I present just below.

- *Fineness of grain*. According to RDApolite, although all SCs share a common response yuk*, not all targets are perceived as contemptible for the same reasons. Some are perceived as contemptible *qua* Jew, other as contemptible *qua* German, and so on and so forth.

The notion "*qua*" (ii) needs to be worked out carefully, but we see already how the categorization component (iii) will provide a sufficient fineness of grain to differentiate among different SCs: if A and B are different categories, being contemptible-*qua*-A is not the same thing as being contemptible-*qua*-B. And for co-extensional SCs, we could add some complexity in either of the components (i) or (iii) of the response.

- *Categorization*. As we saw above, RDApolite allows for Normal token-application of SCs without response. So token-categorization is independent of token-response, hence not driving - nor driven by - it.

But although RDApolite addresses many of the limits of RDAred we identified, it still faces some limitations which will lead us to abandon the project of subsuming SCs under the class of response-dependent concepts altogether.

As we saw in the introduction of the notion of response-dependence, response-dependent accounts of a class of concepts are usually put forward when these concepts are linked to certain metaphysical and epistemological questions. An important limit of RDA is that, in the case of SCs, these metaphysical and epistemological issues are absent, which deprives RDA from a clear motivation. I discuss the two issues successively.

- *Extension*. First of all, response-dependent accounts are usually motivated by a will to escape error theories in cases where realism seems more plausible. As it is unlikely that we are always wrong when thinking about red tomatoes or green leafs, response-dependence makes use of human subjectivity to secure a realist account of colors.

It is partly meant to account for the intuition that it is true that tomatoes are red, facing a red property which is incredibly hard to characterize. Be it for morality or secondary qualities, response-dependence is therefore a natural realist reply to the threat of implausible error theories.

But error theories of SCs are not as implausible as error theories of color or of politeness. It would be odd to claim that nothing in the world is polite or red, but it is not as odd to claim that nobody instantiates slurring concepts. Why would we *want* a realist account of SCs in the first place? One of the main motivations for response-dependent accounts lacks in the case of SCs[138].

The initial motivation of RDA was to account for a real sort of hotness in the concept, but if there are alternative ways to account for the hotness of SCs and the associated expressivity

[138] One remaining motivation could be the will to rescue the coextentionality thesis (and the existence of neutral counterparts), without having to go for a hybrid theory.

of STs, response-dependent accounts are therefore not to be favored. And as we will see in the later chapters, there are alternative ways to account for the hot aspect of SCs.

- *Imunity to error*. Another main common motivation for response-dependent accounts is when there are (sometimes ideal) conditions under which we want to say that subjects could not be wrong in applying the concept.

For instance, assuming that the conditions are ideal and that we are normal human perceivers, it seems meaningless to say that we are wrong in applying GREEN. If the viewing conditions are normal and there is no evil genius and so on, then normal subjects are infallible in their application of color concepts, because token-deployments of color concepts are causally connected to their token-reference.

This observation holds for all other response-dependent concepts as deployed by responders in ideal conditions. By definition, token-deployments cannot fail. Here is a way to understand better why.

It is as meaningless to imagine that normal subjects are prone to error under ideal conditions as it is meaningless to imagine that subjects are wrong about their own experience. Indeed, although we can misperceive something green as something red, we cannot possibly be wrong about the fact that we perceived it as red. Similarly, I can have felt pain without any sort of stimulation of the nerves, but if I felt pain I felt pain and cannot be wrong about that.

Now, assuming that the subjects are normal and that the conditions of perception are ideal, we assume that all potential sources of error are discarded. Every case where the red* response is triggered and the presented object is not red is a case where either the subject is not Normal (his perceptual system could have undergone some change), or the conditions of observation are not Normal (there could be strange lighting, or the subject could wear colored glasses etc.). In virtue of how the red* response works, it is an infallible detector of red things. Whenever it is triggered, it must have been acquainted with a red thing.

The connection is causal, so that every error comes from either noise in the channel or in the receiver. That being said, whether or not the ideal subjects and conditions are in fact possible, it is necessary that under ideal conditions ideal subjects have a perfect red* detector. And since the RED concept is grounded on the red* response, every time a perfect subject has a red* experience under ideal conditions, she can apply RED in an infallible manner.

In a nutshell, Response-dependent concepts use to come with a kind of immunity to error under ideal conditions (see Holton 1991 for more on that issue).

This extends to other response-dependent concepts for which, similarly, it seems that it is meaningless to say that subjects are systematically wrong. A member of the group of polite judges, under perfect conditions of observation, can by definition not be mistaken in her polite* response and hence neither in her token-application of POLITE. That again stems from an inherent connection to non-conceptual responses.

In the case of opaque response-dependent concepts, that connection also has consequences on the acquisition of these concepts. Because one cannot be a Normal possessor unless one has the response, one cannot have acquired the concept without having had the response. This means that opaque response-dependent concept must have been acquired by ostension. Some authors argue that such response-dependent concepts which must have been acquired by ostension are semantically primitive (Wittgenstein et al. 1969, Jackson and Pettit 2002), but that is another issue.

The case of SCs is intuitively very different. We do not need nor want any sort of immunity to error[139]. There is no reason to think that Germanophobes, in virtue of their sensitivity to German people, can hardly be wrong in their application of the concept. And this fact is independent of whether or nor BOCHE has an empty extension.

Assuming the co-extensionality thesis (according to which BOCHE refers to germans), I do not see why we should think that germanophobes can rely on their contempt to detect Germans in an infallible manner. That would be very odd. RDA then lacks another main motivation of response-dependent accounts in that SCs to not seem to require any sort of immunity to error, contrary to secondary quality concepts.

And without the important motivations of rescuing realism and explaining immunity to error, there is no much motivation left for a response-dependent account of slurring concepts, apart

[139] Quite the opposite in fact: subjects are perhaps always wrong in applying BOCHE.

from the close link it introduces between the conceptual and non-conceptual "hot" elements. I thus turn now to that last important motivation.

- *Categorization*. Another common motivation for RD accounts of concepts is when the relevant property is not easily identifiable independently of the response, that is, when the concepts seem *inseparable*:

> **Separability**: A seemingly hybrid concept is separable when its descriptive and attitudinal components are independently characterizable. In particular, a characterization of a separable concept's extension need not invoke the concept's attitude.

Color concepts are inseparable, because it proved very hard to find the necessary and sufficient physical conditions for an object to count as red without recourse to the non-conceptual element (in that case color perception). The simplicity of the human response red* provides us with an easy way to secure the reference of an otherwise highly disjunctive property, and the human response red* thus becomes an inseparable component of the concept RED.

In the case of SCs though, the (intended) extension is very easily identifiable independently of the (i)-(ii)-(iii) response (worthy of yuk*-qua-NC): it is simply the category denoted by the NCs. Given that the third element of the three-layered response is the categorizational component, categorization has taken place independent of the other two elements, a fortiori independent of the hot emotional element. The (intended) extension of BOCHE is the germans, of course.

But if a germanophobes has a "worthy of contempt qua German" response to an individual, she knows her target is German simply in virtue of the "German" part of the response. The contempt did not play a role in categorization, and we can as a consequence identify the (intended) reference of BOCHE independent of the hot element. SCs here become separable concepts: no recourse to human subjectivity seems required to identify their (intended) reference.

But if SCs are separable concepts, the last main motivation of RDA collapses. RDA had that strong advantage to insure a real kind of "inseparable" hotness to SCs. But having been led to introduce a three-layered response including an element of cold categorization (iii), the hotness of SCs is not essential to SCs in the sense it was initially meant to be.

If we become open to conceive of SCs as separable concepts, there are plenty of alternative ways to account for hotness, and if we can find another satisfactory way to account for hotness, there is not much motivation left to prefer a response-dependent account over another kind of account.

I will now evoke a possible reply that a proponent of RDA could give to address the present issue of separability. Although it is true that by introducing a descriptive bit (iii) in the response itself, a kind of separability was reintroduced in the concept. The descriptive bit needed not be the neutral counterpart, it could instead be a stereotype for instance.

Again, a stereotype is a set of properties that the (intended) reference is taken to instantiate. The stereotype for BOCHE could count properties such as cruelty, strictness, obedience for instance. Under that analysis, the relevant response in BOCHE would not be "worthy of yuk*-*qua*-German" anymore, but rather "worthy of yuk*-*qua*-cruel/strict/obedient…". This gives us a new version of RDApolite:

> **RDApolite'**: it is (Normally) *a priori* that x is a SC iff x would cause NC-phobic people to have a worthy-of-yuk*(i)-*qua*(ii)-Stereotype-for-NC(iii) response under Normal conditions of slurring thought.

RDApolite' addresses the issue of separability, because stereotypes do not fix the reference (according to Putnam). More, a stereotype is somewhat descriptive in the sense that it is a set of properties, but it is not separable from evaluation. When trying to make the stereotype explicit, we seem forced to use thick terms like "cruel" or "lazy" that are themselves inseparable.

But there is a rebuttal of this answer to the question of separability. Indeed, even though the stereotype itself must be described with thick inseparable terms and is therefore not independen from an evaluation, it is independent from the evaluation which is relevant in the SC.

The existence of the term "Jewy" is a good illustration of this fact. Whether or not the stereotype is itself evaluative, it seems that "Jewy" refers to people who satisfy the stereotype of Jewish people, whether or not they are actually Jewish. The term thus targets people who satisfy a stereotype, whether or not they have the property denoted by the NC.

For the existence of such a term to be possible, the conceptual/referential element in KIKE must be separable from the stereotypical/evaluative element. If the two were inseparable, how could speakers have come up with a term keeping only one dimension, the stereotypical dimension? "Jewy" could not exist in the language if "kike" was inseparable, it seems.

So the answer just considered is off mark: replacing the descriptive NC component of the response with a descriptive Stereotype component does not really escape separability. We will discuss other limits of the recourse to stereotypes in the next chapter.

- *Opacity*. If SCs are in fact response-dependent, it is not self-evident. The subjects who deploy BOCHE and use "boche" hardly know that the category might depend on their own responses. They most likely use/deploy slurring representations to refer to germans, and might in some cases not even notice that the concept is loaded with their own negative responses in one way or another.

So if SCs were response-dependent, we would most likely be in an opaque case of response-dependence rather than in a transparent one. But POLITE is not opaque, and we saw that if SCs are opaque like RED, we run into a series of problems having to do with recognition, categorization, possession conditions, extension etc.

Overall, because of our understanding of concept individuation, response-dependent accounts of SCs lack a clear motivation. SCs do not seem to require a non-empty extension (which might be a requirement), they do not come with an intuitive immunity to error, and their hotness cannot be characterized in the strong inseparable way we wanted.

Coming back to earlier observations, I want to argue that a response-dependent account of SCs might be true, but only under a very general understanding of response-dependence which does not enlighten the question of SCs as it would equally apply to all empirical concepts. In the sense that all empirical concepts are grounded on a canonical mode of presentation, it is true that normal concept possession requires having the response. Normal subjects who deploy the concept can thus be taken to be responders, and so on and so forth.

But that general kind of dependence to a canonical mode of presentation, that kind of response-dependence, applies to PEANUT and CROISSANT. And a theory that does not distinguish PEANUT from BOCHE cannot be a satisfactory theory of SCs in their specificity and particular features. Something additional distinguishing SCs from PEANUT, on top of their being response-dependent, must be introduced. We will see that an essential component in SCs might play a non-negligible role in distinguishing them from other empirical concepts.

We initially concluded after the discussion around RDAred that a more substantial notion of response dependence was needed to account for SCs. But we now see that even the more substantial versions fail. Either we draw too strong an analogy with secondary quality concepts to be able to explain the actual behavior of SC possessors, or we loosen the analogy by allowing transparency and hybrid responses, and RDA ends up drained of its promising original motivations.

Chapter 10. An Essentialist Account

In this chapter, I will put forward an essentialist understanding of SCs. I will start with a small introduction on the notion of psychological essentialism, and show that it is intuitively appealing in the case of SCs.

I will then consider successively two models of SCs as essentialist concepts. The first is based on a modal conception of essences put forward by Putnam (and Kripke), according to which essences are necessary properties. An object cannot exist as an "O" without having the essence of an "O".

I will briefly consider the advantages of such a view applied to SCs (EAputnam), and see that it faces limitations that I will discuss. One objection to the first view will be that it is possible to be ascribed a SC even without having any of the negative essential properties that slurrers take their target to have. I will answer that what is necessary is the deep, invisible essence itself, not the negative stereotypical ensuing properties.

A second objection will note that even the deep essence itself is sometimes seen as unnecessary for the target to fall under the SC. This objection is based on a example put forward by Jeshion (p.c.), and I will consider four possible ways to respond to the objection and concede that the cases at hand weaken our first essentialist view of SCs.

This is why I will put forward another version of an essentialist account of SCs based on an Aristotelian understanding of essences, seen not as necessary properties but more as something like an underlying "nature". In an Aristotelian model of essences, "monstruous" cases, where the thing exists without having the essence, are *possible*, which resolves many of our earlier issues.

I will see how such a view applied to SCs (EAaristotle) handles the main explananda we started with. The notion of a negative essence, that is of an essence which is value-laden, will play a non-negligible role in accounting for some explananda and will hence be given a closer look.

10.1. Essences

If slurring concepts are not response-dependent concepts, what kind of concepts are they? To pinpoint other theoretical options, consider the following Dialogue (from Richard Powers, *The Time of Our Singing.)*.

> *[During segregation, Dr Daley is rejected from a symposium because he is black]* - You see, I couldn't in fact be Dr William Daley of Philadelphia because Dr. Daley is a real medical doctor with genuine credentials, while I'm just a nigger busting his wooly head into a civilized meeting of medical professionals. [...]
>
> *[David a jewish physicist gives the following explanation]* - What has been done to you today. This is an error of statistics. ... They are taking shortcuts in the steps of their deductions. They did not see the case, but only made bets on the basis of what they think likelihood tells them. Category. This is how thought proceeds. [...]
>
> *[Dr Daley does not accept this justification and answers]* - Likelihood be hanged. This is nothing but animal hatred. Two species. That's what they see. That's what they're intent on making. [...]

Two views of SCs are opposed in the dialogue: David's view is statistical/stereotypical, and Dr Daley's view can be coined essentialist. Both have an intuitive appeal and deserve being developed and assessed. Dr Daley mentions the notion of "species" to oppose David's statistical account. There is a strong intuitive sense in which Dr Daley is after something deeper and more accurate than David. My Essentialist Account (EA) is an attempt at making that intuition explicit.

We saw that the major issues of stereotypical or statistical accounts of SCs had to do with the need to postulate a property deeper than any (combination) of the surface stereotypical properties. Drawing upon the notion of essence, EA preserves the explanatory powers of the stereotype, but introduces an additional deep causal source for that stereotype. This additional and more important ingredient in SCs is, as we shall see, suited to account for all the main explananda we identified.

Importantly, what really matters for categorization as a member of a class is the deep essence, not the surface stereotype. Stereotypical properties might be used as a clue that there is a hidden essence, but these properties would play only a derivative or secondary role. Only the hidden essence is decisive for membership in the class.

Under this intuitive version of the view, one can still be a member of the class without having any of the stereotypical properties, which seems like a good start to address the main problems of the stereotype view.

SCs could thus be essentialist concepts. But as there are many essentialist concepts, that claim is not substantial enough to characterize the specificity of SCs. Admittedly SCs postulate an essence, but what kind of essence exactly? Remind that the apparent specificity of SCs is that they seem to have a hot component, an inherently evaluative bit. We could thus be tempted to claim that SCs ascribe not any kind of essence to their targets, but an essence that is loaded with a value. We should thus make some room in theorizing for the attribution of a valence to essences themselves.

EA could thus be formulated in the following two tenets, that we will investigate in this chapter:

> **EA**: *i)* SCs are essentialist concepts; *ii)* SCs attribute *negative* essences.

The first tenet of EA assimilates slurring concepts to essentialist concepts, that is, it claims that SCs ascribe an essence to their targets. I dig into the notion of essence below to understand better what that claims consists in and the explananda it is able to take care of.

The second tenet claims that the essence which is ascribed by SCs to their targets is value laden. The essence is not taken to be a neutral element that causes potentially valued surface properties; rather, it is itself loaded with value. I will discuss the different ways to flesh out this idea below.

Overall, the two simple assumptions in EA, according to which SCs wrongfully ascribe negative essences to their targets, enable us to derive most of the main properties of SCs, given independently plausible assumptions about negative essentialist thinking. Let us thus start with a brief presentation of essentialism.

No characterization of essentialism is unanimously accepted. Essentialism is roughly the view that some objects have *essential* properties. We will define the notion of essential property below, but to give a rough idea, the essence of an object is the underlying nature that makes it the thing it is. The essence of a cat could be the cat's DNA for instance.

Essentialism is thus a metaphysical claim about the nature of reality: it says that at least some things have essences. But I need not take a stand on the nature of reality to describe slurring concepts. Whether there really are essential properties does not really matter for the present investigation, as there might well be essentialist *concepts* without there being essential properties:

> One can believe that a category possesses an essence without knowing what the essence is (Medin & Ortony, 1989)

In other words, some concepts could simply postulate that there is an essence in the objects they apply to, even if there is no such essence. This possibility is all that matters in the present exploration of the view that SCs are essentialist concepts. *Psychological* essentialism is the view that human naive psychology is basically essentialist, that is, that most of our natural concepts postulate an essence in the objects they apply to.

There are two main families of views on the nature of essences. Roughly, the first identifies the essence of an object with a property that the object necessarily possesses. I call this view "Putnam's view of essences".

The second identifies essences not with a property, but rather with an unanalyzable nature of the object, which is the causal source of the important properties that the object has. I call this view "Aristotle's view of essences".

I now examine successively the two conceptions of essences and consider their respective explanatory power when applied to SCs.

10.2. Putnam's View of Essences

The standard conception of an essential property, which can be attributed to Putnam and Kripke, is that of a property that an object necessarily possesses at all worlds. Based on this modal conception of essences, consider a first definition of essences:

> **Essence1**: Something is an essence iff it is a property that an object necessarily possesses at all worlds.

Under this modal conception of essences, a property will count as an essential property of an object or individual just in case that object or individual possesses the property in all other possible worlds (where it exists). As a consequence, an object can in principle have many different essences.

For instance, being constituted of H2O will count as an essential property of water, for in all worlds such as twin-earth where something has all the properties of water except that it is not constituted of H2O, the thing in question cannot count as water, and so even if it possesses all the usual surface properties of water such as being transparent, odorless etc.

A property that is not essential is said to be *accidental*. An accidental property is such that it is *possible* that an object exists while lacking that property. For instance, water can lack the property of transparency and still be water, (that is still be composed of H2O), when it is frozen for instance. Transparency is thus an accidental property of water, whereas being composed of H2O is an essential property.

Applying this modal conception of essences to EA would give us the following version of the view:

> **EAputnam**: i) SCs attribute necessary properties to their targets; ii) The necessary properties that SCs attribute to their targets are negative.

According to this view, a concept like BOCHE attributes a negative necessary property to German people, that is, a property that German people cannot possibly lack and that is negatively valued. Such a necessary property is (what is taken to be) the essence of German people.

But before looking at how EAputnam would account for our explananda, we should focus on a first simple objection: it does not seem *prima facie* that there are any necessary property or set of properties associated with group-membership in the case of slurring concepts.

10.2.1. Objection 1: Unnecessary Properties

We can draw the following simple objection to EAputnam[140]. Many negative properties associated with a SC are not (deemed to be) metaphysically necessary. For instance, someone categorized as a "boche" is not always judged to necessarily be cruel, etc. More generally, it seems hard to find any alleged negative property of SC targets that is taken to be necessary.

In fact, as we saw already when considering the stereotype view, it is always possible/conceivable to fall under the (intended) extension of a SC without being taken to have (any of) the negative properties of the class. So even though an essentialist account of SCs is appealing, there is no likely candidate to play the role of the negative essence for a given SC.

A possible reply is to distinguish the essence itself (the property of "bocheness", whatever that is supposed to be) and the stereotype, which is its superficial (i. e. more epistemically accessible) manifestation. Stereotypical qualities, though taken to somehow derive from or indicate the underlying essence, need not be metaphysically necessary, on a Putnamian view.

It is thus important to distinguish the essence itself, which is a necessary property, from what we could call *essential properties*, which are those of the accidental properties that derive from the essence.

The essence is necessarily instantiated by essence-holders, so long as they exist. But the properties that derive from the essence are not necessarily instantiated. For instance, cats are cats in virtue of having a certain essence (say the DNA of a cat), and their essential property

[140] Thanks to Robin Jeshion for this objection (p. c.)

of having a tail derives from this essence. But a cat could have no tail for one reason or another and still be a cat. Having a tail is an *essential property* of a cat, and having the DNA of a cat is the cat's *essence*.

Applying that distinction to SCs under the scope of EAputnam allows us to restore a certain role to stereotypes, but without giving them too much. Stereotypes would be identified with essential properties (see below). Germanophobes would attribute a negative essence to german people, and take a set of surface properties - such as cruelty or obedience - to be derivative from that negative essence.

Overall, when talking about essences, we should at least distinguish between the following levels:

> **Essences**: An essence is whatever is responsible for *i)* kind membership and *ii)* an array of observable features (be it metaphysically real or not). The essence of a cat could be it's DNA for instance. There is room for physically instantiated essences such as DNA, also for historical or social essences.
>
> **Essential properties**: Essential properties are (maybe causally) derived from the essence. Subjects tend to rely on the detection of essential properties to infer the presence of an underlying essence. The cat's claws or its fur could be such essential properties for instance. The essential properties associated with SCs could correspond to the stereotype.

And it is useful to add the following sub-distinctions:

> **Derivative essential properties**: Derivative essential properties are not directly derived from the essence, but they accompany essential properties. They roughly correspond to what evolutionary biologists call "spandrel". For instance, the eye's blind spot is a consequence of human DNA (hence of the essence), but it was not really selected for its advantages in fitness, it is rather a necessary bad accompanying the eye. The cat's blind spot thus belongs to its derivative essential property, whereas its fur or claws are real essential properties.

> **Accidental surface properties**: Accidental surface properties are observable and not connected to the essential property in any way. They do not play a role in categorization. The fact that cats are made out of a lot of water (like all living things) constitutes an accidental surface property for instance.

Clearly distinguishing between these levels, and especially between the two levels of the essence with a (possibly negative) necessary property and the stereotype as a set of essential properties, allows us to escape Jeshion's objection that there are no observable necessary property associated with group-membership in the case of SCs.

Indeed, we can now better conceive of the possibility of being ascribed an SC without having any of the essential properties: no property appears to be necessary precisely because the only necessary property that is crucial for categorization - the essence - is by definition a hidden property.

Importantly, this objection applies to Neufeld's (2018) view of slurring concepts as essentialist concepts causally connecting the essence with negative surface stereotypes. Without the essence itself being negative, we loose both the ability to explain slurs with only positive stereotypes, and as we shall see a neat account of the dehumanizing power of slurring representations.

10.2.2. Objection 2: Unnecessary Essence

But there is another, more powerful objection to EAputnam that a distinction between *essences* and *essential properties* is unable to respond to. This objection is not fatal to EAputnam, as we will see with the five tentative replies I give below, but I argue that it provides sufficient motivation for an independently plausible alternative version of EA based on an Aristotelian conception of essences.

The objection is an elaboration on Jeshion's first objection and notes that, in many cases, even the essence itself (as opposed to any surface properties), is not judged to be metaphysically necessary.

We can observe this fact in examples like the following. According to EA, FAGGOT applies to those who have the "gay" essence, whatever that may be, and this essence is responsible for their stereotypical surface properties, whatever they may be. But when homophobes conceive of Harvey as a FAGGOT, they do not usually seem to judge that Harvey could not have existed without being a "faggot". The essence ascribed by homophobes to homosexual people does not seem to be necessary, because it seems that the existence of the individual would still be possible without it.

Even more, when calling Harvey a "faggot", it seems that the homophobe condemns Harvey and judges that he would be better off not being so. But if homophobes believe that homosexuals ought not to be homosexuals, they must consider that it is possible for them to not be homosexual.

Arguably judging that someone oughts to not have a property P presupposes that existence is possible without this property. P is here taken to be metaphysically unnecessary. But according to EAputnam, the ascribed essence is supposed to be metaphysically necessary.

This objection is stronger than the previous one. I propose the five following lines of reply, before granting the point and modifying our current conception of essences.

My first reply to this objection is simply that the intuition it relies on - the intuition that homophobes judge that homosexuals should be heterosexual - is not so clear. Of course, some homophobes seem to believe that "homosexuality is a choice" or "homosexuality is a lifestyle", thereby implying that there is an alternative and that homosexual's homosexuality is then not necessary, nor essential.

But homophobes who say that homosexuality "is a choice" can in fact still be essentialists. They could believe that there is an essence of "faggot" that their target instantiates, but that the people who have it ought to better control the *expression* of this negative essence.

We should not conclude from the homophobes saying that homosexuality is a choice or a lifestyle, or that homosexuals should be heterosexuals, that the concepts they use to slur their target is not essential. Homophobes could as well believe that their targets possess a negative essence of "faggot", but should strive to keep it quiet and not express it in having same sex relationships. In other words, maybe that homophobes do as if it was necessary that the target is a "faggot", but as if it was accidental that he has the surface properties.

My second worry is that the objection might apply to only a narrow subclass of SCs. It might be the case of some slurring representations like "faggot" that it does not postulate an essence, whereas other slurring representations, maybe racial ones, do. Targeting people because of their origins is not the same as targeting people because of their behavior.

By way of a third reply to the objection, I want to remark that there are different levels of essences. We should distinguish between *sortal* essentialism and *origin* essentialism. Under a sortal conception of essences, I am a human being because of a property of being human that is essential to me. Under an origin conception of essentialism, my biological origins - that is the sperm and egg from which I arose - are essential to me.

Under a sortal conception of essences, one can exist as one and the same individual with a modification of essence. It is in virtue of this mechanism that a tadpole can become a toad and still be the same individual. It was a tadpole in virtue of an essential property of being a tadpole, and it is now a toad in virtue of another essential property of being a toad. Such changes of essences are not conceivable under an origin conception of essentialism, because the tadpole and the toad have one and the same biological origin.

This distinction is telling in the case of SCs when it comes to the second objection, because homophobes might have taken "faggots" to have a sortal essence that they can acquire or loose, rather than an origin essence that is intangible. Hence their conceptions of homosexuality as a choice or a way of living, and their judgements that homosexuals ought not be so, still under an essentialist understanding of homosexuality.

Fourth, it should be noted that the objection considered presupposes that "ought" implies "can", which is not necessarily the case (see e. g. Saka 2000). So the fact that homophobes believe that "homosexuals ought not to be homosexual" does not entail that they take their homosexuality to be non-necessary and hence non-essential.

And finally, even if "ought" implies "can", homophobes could well impose a double bind on homosexuals when thinking that they are essentially homosexual but ought to be heterosexual. SCs would attribute an essence to individuals and condemn it at the same time, supposing that targets should strive to loose their essence, which is impossible. But ascribing an impossible task to their targets shall not necessarily discourage homophobes to be homophobes and apply FAGOOT to homosexuals.

Based on the five previous replies, it seems that subtler versions of EAputnam could perhaps reply to the second objection. But let us grant the point, and admit that some cases do behave as Jeshion notices - that there are essences and natures we ascribe without taking them to be necessary.

There is still some theoretical space left for EA though. Let us consider another model of essences taking the non-necessity of essences into consideration. Instead of a modal conception of essence, we can take a notion of essences which is closer from the one that is used in the psychology literature.

10.3. Aristotle's View of Essences

An alternative conception of essences can be traced back to Aristotle, is used in the psychology literature and addresses some issues with the modal conception of essential properties, *à la* Putnam.

One of the main issues with merely modal conceptions of essences is that it does not account for the conceivability of properties that an object possesses at all worlds, and which are not the underlying cause of surface properties. For instance, at all worlds where cats exist, cats have a shape, but their having a shape is not what constitutes their being a cat, for anything with a shape would otherwise be a cat.

So on top of being necessary, essences must be causally linked to the relevant syndrome. In other terms, an additional layer is needed: there are essential properties and surface properties, but among surface properties, there are those that are essential in the sense that they are caused by the essence, and those that are accidental because they are causally disconnected from the essence.

This conception is closer to what Aristotle seemed to have in mind, and to the current usage in the psychology literature. That is the sense of Medin and Ortony's second tenet:

> First, people act as if their concepts contain essence placeholders that are filled with "theories" about what the corresponding entities are. Second, these theories often provide or embody causal linkages to more superficial properties. Our third tenet is that organisms have evolved in such a way that their perceptual (and conceptual) systems are sensitive to just those kinds of similarity that lead them toward the deeper and more central properties. (Medin and Ortony 1989, p. 186)

These considerations allow us to consider a second notion of essence which could be applied to SCs:

> **Essence2**: An essence is the nature of a thing or kind of thing. The essential properties of a thing are those that derive (in the right way) from its essence. Crucially, a thing can exist without manifesting its essential properties. For example,

> having two arms is an essential property of a human (it is part of its nature). Accidentally, however, a human may have less or more than two arms.

Essence2 differs from Essence1 in important respects that will help us address the problem of unnecessary essences raised against EAputnam. Under an Aristotelian conception of essences as a defining nature of the being having the essence, the essence is not necessarily something that the being has in all the worlds where it exists. The model only characterizes the essence as the nature of the object, but there is some room left for the object being what it is without having the essence.

For instance, it is an essential property of a cat that it meows, because it is a consequence of it's underlying nature. But since there are cats which happen not to meow (because of diseases or malformations for instance), under a Putnamian conception of essences, meowing cannot be an essential property of cats. Mewing would then be an accidental property. In other terms, there are certain properties that would be considered as accidental under a Putnamian framework that are considered essential under a Aristotelian framework.

This is an advantage of Essence2 over Essence1, because there is intuitively an important difference between properties like meowing and properties like being dirty for a cat, whereas a modal conception does not distinguish between the two kinds of properties.

Aristotle's conception of essences leaves some room for "monstruous" cases where an object fails to have an essential property but still falls under the concept. The object would still be taken by the concept possessors to have the essence, but there would have been a failure in the process of implementation of the essence. The connection between the stereotypical syndrome and the possession of an essential property is then more corelational than causal.

I argue that this conception of essences is more appropriate to SCs. Let us then examine an improved version of EA:

> **EAaristotle**: i) SCs attribute an underlying normal cause to the syndrome of their target's; ii) This cause is an essence and has a negative valence.

If SCs are essential concepts in the sense defined by EAaristotle, the problem of unnecessary essences can be addressed. Homophobes would believe that their target's nature is that they are "faggots", believe that "faggots" ought not to be such, and if ought implies can, believe that it is possible for "faggots" to loose their essences.

Under this story of the case at hand, homophobes who believe that homosexuals should not be homosexuals do not only want that their targets modify their apparent behavior, but they ask for some sort of a conversion: a change in their very nature.

Let us now focus on the evaluative component of the essence. The specificity of SCs, among essential concepts, would be that the essence they postulate is in itself negative. But what is it exactly for an essence to be negative? I see at least two ways for an essence, understood as the underlying cause of a syndrome, to be negative.

Either it is the essence itself that is negatively loaded, independently of its power to cause negative essential properties, or the essence is negative only inasmuch as it tends to cause negative essential properties. Under the first view of the evaluative component, an essence can be negative even if none of its manifestations is negative. This could account for the SCs associated with positive stereotypes.

For instance, some sexists may make only positive statements about their targets, such as: "woman are more elegant and intelligent than men". In doing so, they would ascribe only positive surface properties to their targets. But they still deploy an essentialist way of thinking, and in this sense betray their hidden belief that woman have an essence of women, potentially a negative one which happens to cause only positive properties such as intelligence and beauty.

These forms of sexism would not be possible under the second conception of the essence's value. If the value of the essence comes from its propension to cause negative/positive surface properties, then only targets associated with negative stereotypes could be referred to through the deployment of a SC. This is why I think that the first analysis of the evaluative component - according to which the essence itself is taken to be negative, independent of the valence of its manifestations - is preferable.

SCs end up having an interesting property, which makes it superior to a purely stereotypical view that David was putting forward in the initial dialogue. First, in a world where all members of the target class happen to accidentally have the stereotypical features, the Slurring concept stays flawed. This would not be the case of a SC understood as a stereotype, because if it so happened that the target met the stereotype by accident, SCs would simply be appropriate.

There might be a fruitful connection to draw with current debates on the "true self". There is allegedly an asymmetry between good and bad properties with regard to the true self. Psychologists indeed claim that subjects of experiments judge that loosing a bad property makes us closer to ones "true self" than acquiring one (Newman, Bloom, & Knobe, 2014).

For instance, being addicted to a substance is usually perceived as a bad property. A former drug addict having ceased to be addicted to drug is hence perceived as now closer to her true self than before she stopped being addicted to drug. On the opposite, someone becoming a drug addict is usually perceived as moving away from her true self. So when it comes to the self, it seems that bad properties are accidental while good properties are essential.

The opposite is the case of SCs, under EAaristotle. According to the present view, SCs ascribe negative essences to their target. Consequently, the bad properties are essential and the good ones are accidental. SCs would thus function like concepts like KILLER to that matter. When one applies such a concept to an individual, it seems that we do more than simply stating that the target killed someone.

It seems that being a killer is part of the target's nature, or something along those lines. And being a killer is a bad nature, of course. SCs, on a par with concepts like KILLER, ascribe negative essences to their targets, as if the true self of the targets was negative, whereas the true self of human beings, as we saw, is positive. Hence the dehumanizing power of SCs is understandable, as we shall now see.

Let us see how EAaristotle accounts for the major explananda we identified.

Let us now consider how EAaristotle deals with the major explananda for SCs.

- *Hotness*. According to the second tenet of EAaristotle, essences are value-laden. The hotness of SCs could thus simply derive from that of value, whose "hotness" should be independently accounted for.

- *Defectiveness*. Each of the two tenets of EAaristotle can be responsible for the defectiveness of SCs, one cognitive and another moral/ethical. First, SCs are defective because there is no such thing as the essence of a group of individuals. The corresponding essences do not exist, and SC therefore encapsulate a misrepresentation of their targets.

People who deploy SCs are thus wrongfully essentializing their targets, and ascribing an essence to social groups misrepresents the social realm. They assume that some observed surface properties of groups of individuals are caused by the presence of an underlying property, a bit like the DNA of cats is the usual cause of their stripes.

Importantly, this embeds a terrible mistake, as there are no social kinds. Although essentialist thinking is well fit for natural kinds, it is misplaced in the social realm. Such a misconception is in itself the source of social injustice and harm. The first tenet of EA is thus responsible for a cognitive flaw: a false belief about social essences[141].

So essentializing is one of SCs cognitive flaws, but it is not the only one, nor is it the most characteristic defect of SCs. Neutral terms can be wrongfully essentializing too. Consider again the sexist and essentializing remark "Women are more clever than men". Such a remark seems to encapsulate the same harmful kind of essentializing mode of thought as

[141] Focusing on the case of race, there is an existing debate. Racial anti-realists (i. e. Glasgow 2009, Zack 2002, Blum 2002, Appiah 1996) argue that there is no such thing as a race. Social constructivists (e. g. Haslanger 2008, Taylor 2013, Sundstrom 2002, Root 2000) argue that "races" exist as a social construct. Biological racial realists (e. g. Risch et al. 2002, Mayr 2002, Kitcher 1999, Andreasen 1998) argue that the notion of race has, to some extent, some biological foundation.

slurring representations targeting women, although it does not involve a slurring representation.

The second flaw of SCs, which also seems to be more specific to them, comes from the fact that the value associated with the alleged essence is negative. That is, on top of falsely believing that there are essences for social groups, NC-phobic people ascribe an inherently negative value to their target's identity. Here stands an additional source of all the wrong attitudes and actions of NC-phobic people. The second tenet of EA is thus responsible for a moral/ethical/political flaw: they wrongfully ascribe an inherently negative value to a social entity.

- *Extension*. EAaristotle predicts that SCs have null extension. There are no individuals who possess the essence of a "boche", and hence no one to fall under the extension of BOCHE.

Although SCs have a null extension, they have an identifiable *target*. We should thus carefully stress our earlier distinction between *target* and *reference*. SCs target the reference of NCs, but fail to refer because of the false presupposition they carry[142].

This distinction between a reference and a target is reminiscent of Donnellan's (1966) distinction between referential and attributive uses of descriptions. When someone at a party asks "who is the man drinking a martini?", wondering about a man who is in fact drinking a glass of water, the description "the man drinking a martini" fails to refer because of the false presupposition that the man is drinking a martini.

But the description can arguably be said to have a target though, because the speaker intends to refer, in fact, to the man drinking water. The kind of failure involved in SCs understood under EAaristotle is similar to this kind of failure. Their extension is null, but their target salient.

- *Possession conditions*. Under EAaristotle, one condition to count as a Normal possessor of a SC is that one ought to attribute negative essences to the targets. Since SCs are essentialist concepts, a speaker who is not (psychologically) essentialist about the target, say about Germans, cannot be said to really possess the concept of BOCHE. So Normal possessors attribute negative essences to their targets.

[142] On the moral flaw of postulating essences, see Leslie (2013, 2015).

What about parasitic possessors, that is, about people who appear to understand racists when they are using STs, and seem disposed to deploy SCs themselves? One possible answer could be that parasitic possessors of SCs use concepts similar to SCs, but without the essentialist component. The difference between real essentialist SCs and their non-essentialist counterparts as possessed by non-racists should be fleshed out in detail.

- *Dehumanization*. EA is very well equipped to offer an original explanation of the important dehumanizing power of SCs. Indeed, ascribing a negative essence to human beings contradicts the otherwise plausible assumption that human essence is positive.

We discussed above two possible ways for an essence to be negative: the essence itself could be negative or it could be negative only inasmuch as it is causally connected to negative properties. This difference between negative essences and negative essential properties can be best seen under the light of Darwall's (1977) distinction between *recognition respect* and *appraisal respect*.

There are according to Darwall two kinds of respect that we can demonstrate towards human beings. One kind of respect (appraisal respect) is grounded on the manifestation of special qualities which make the person worthy of positive appraisal. We can respect someone *as a* philosopher for instance, and that would be an instance of appraisal respect.

Another kind of respect

> "consists in giving appropriate consideration or recognition to some feature of its object in deliberating about what to do" (Darwall 1977, p. 38).

This sort of respect is precisely the "sort of respect which is said to be owed to all persons". That all persons are worthy of recognition respect *qua* persons, not in virtue of their qualities, means in essentialist terms that the essence of all persons is positive. Here is the flaw of SCs.

With the notion of recognition respect at hand, we can hypothesize that what goes wrong in SCs is not only that it falsly ascribes an essence to the targets, but that on top of that, because it supposes that the essence is negatively valued, it fails to recognize the dignity of their targets as human beings with a positive essence. There is indeed a conception of dignity following directly from human nature, and ascribing a negative essence to a human being is failing to recognize the person's dignity as a human. Hence it is excluding it from the group of humans, it is dehumanizing.

A good prediction of this view is that targets categorized as SCs can be attributed overwhelmingly positive qualities. Yet SCs are still dehumanizing, just like in my earlier example of a sexist believing that "woman are pretty and intelligent and wise". Dehumanization comes from the rejection of human dignity inherent to the deployment of SCs, not from the ascription of negative qualities.

- *Identifying thinking*. This is definitional of essentialist thinking: the target's identity is reduced to the essence she is taken to have. Being a "boche" is what the target *is*: it is its nature, its essence.

- *Ideologies and Stereotypes*. We saw that SCs seemed to be associated with racist ideologies and stereotypes about the targeted groups. Ideologies can be seen as explanatory frameworks. In the case of SCs, a "slurring ideology" would be one that appeals to negative essentialist (pseudo-)explanations of social phenomena. More, essences are also associated with stereotypes.

In the Aristotelian framework, stereotypes include both stably associated accidental properties, and essential properties (that is, properties ensuing from the underlying essence). But contrary to a purely stereotypical view of SCs, where SCs would be *identified* with the stereotype, SCs do not merely amount to a statistical error, as we saw earlier.

- *Neutral Counterparts*. NCs may or may not be essentialist, but in any case, they do not postulate not *negative* essences. Under EAaristotle, there is room for a conception of NCs as associated with purely negative properties, but not yet equivalent to a SC, because the essence *itself* would not be negative.

- *Contempt*. We introduced the idea of value-laden essences. In particular, the essence postulated by SCs would be of a negative value. Now, there might be such a thing as a "fitting" attitude towards essences. We saw earlier that "recognition respect" was, according to Darwall (1977), the "fitting" attitude towards human beings. A lack of recognition respect towards a human being hence constitutes a moral impairment, because every human, as a human, is worthy of recognition respect.

Similarily, "contempt" could be the fitting attitude for negative-essence-holders. Were anyone really instantiating such a negative nature, contempt would be the appropriate reaction to this individual. Therefore, possessors of SCs feel something like contempt towards their target precisely because they (wrongly) take them to instantiate negative essences.

Interestingly, this view differs from RDA in that the emotional reaction of contempt arrives after the cognitive act (here, the postulate of a negative essence in the target). RDA tried to derive the structure and function of SCs from an original automatic reaction of contempt. EA reverses the order. In some sense then, the cognitive flaw precedes the moral one.

- *Projection/Offense*. Recall the Use-Possession rule' (UP') I introduced earlier:

> **Use-Possession rule' (UP')**: Language users usually possess the SC that the ST they use express.

If such a rule is at play in conversation, then hearers infer possession from use. Now that SCs are taken to be negative essentialist concepts, hearers can also infer from the possession of a SC that the possessor has a negative essentialist concept about her target from possession.

This move gives us at the same time projection and offense. It gives us projection because the mere fact of possessing or deploying a SC, be it embedded or not, shows that the speaker is thinking about her target in a negative essentialist manner. And thinking about individuals as if they instantiated negative essences is offensive, because it is dehumanizing.

- *Derogatory autonomy*. A potentially interesting consequence of EAaristotle is that it will lead to deny derogatory autonomy. Hom (2008) has argued that the derogatory force of STs was autonomous from the beliefs, attitudes, or intentions of their users. The same could apply to SCs, as we saw. Now, under EAaristotle, what is derogatory in SCs is precisely that their possessors, hence the users of STs, *believe* that their target has a negative essence. The derogatory force of SCs is then not autonomous from the beliefs or attitudes of their possessors, quite the opposite.

- *Derogatory variation*. I see at least three possible factors which could explain derogatory variation, that is, the fact that some STs are more offensive than others. First, as soon as we have an essence and properties deriving from the essence, some coextensional SCs could impose some sort of a "filter" on the negative properties.

For instance, NIGGER and SPADE would postulate the same negative essence in their target, but the former would direct attention toward a certain set of negative essential properties (e. g. being lazy), whereas the later would direct the attention toward another set of essential properties (e. g. being "cool"). This notion of a "filter" on negative stereotypical properties could account for the inter-group derogatory variation.

Second, we saw that EAaristotle proposes to ascribe a negative value to essences themselves. Now, depending on the nature of the negative value, the negative value of essences might very well itself come in degree.

Third, two distinct SCs could ascribe a negative essence to their respective targets, but one could be taken to instantiate more negative essential properties than the other. One SC could come with a negative essence which itself causes many diverse negative essential properties (the stereotype), whereas another SC could come with a negative essence which, for some reason or another, does not lead to as many negative essential properties.

This could be the case of the sexist believing that "women are more intelligent than men" for instance. He would still ascribe a negative essential property to his target, but in this particular case, his ideology comes with something that prevents the negative essence from becoming manifest in negative essential properties. The stereotype thus becomes positive. As a result, one ST associated with a positive stereotype might end up being perceived as slightly less derogatory than another associated with a negative stereotype.

General Conclusion

The present dissertation was aimed at better understanding slurs, their structure, their function(s), their cognitive underpinnings, and the theoretical lessons we could draw from their existence in natural language.

I came up with a set of explananda with regard to which we could evaluate the different theoretical accounts of slurs. We saw that properties such as projection and expressivity were particularly important.

I investigated the notion of hybridity, and compared slurs with other expressions and constructions in natural languages displaying a similar sort of "hybrid" content. In particular, we saw that presuppositions, conventional implicatures, and conversational implicatures constituted promising bases for a hybrid linguistic account of slurring expressions. I have therefore successively explored, and argued against, each of the three views.

I have then put forward a general objection to each of these three linguistic views. The objection is based on the observation that hybrid accounts, even though they are descriptively adequate (as the conventional implicature account might be), lack a clear theoretical framework. Describing a linguistic phenomenon is one thing, explaining it is another. I have argued that an important dimension has been neglected in the debates surrounding slurs: psychology.

So I made the bet that slurs derive most of the interesting property they have from features of a mental representation, a concept, that they are used to express. I have then dedicated the remaining of the dissertation to pursue an account of what I have coined "slurring concepts".

I have attempted at building a first view of slurring concepts by questioning one of our starting hypotheses: the hypothesis that slurring terms and concepts are *hybrid*, that is, that their semantics contains two dimensions. The resulting view consists in treating thick terms and concepts as, appearances notwithstanding, not truly evaluative: they simply have a rich descriptive content such that they refer to subgroups, subgroups which are independently, extra-semantically evaluated as being negative. The evaluation ends up being associated with the terms, but it becomes associated only extra-semantically. The semantics of these terms would thus be one-dimensional.

I have then explored a more radical theory of slurring concepts locating all of their dimensions, including the evaluation itself, in the truth-conditional layer. According to such a view, that I have called the "truth-conditional account", slurring concepts would simply be complex descriptions such as "worthy of contempt because of...". I have argued that such a view would have a hard time accounting for projection facts. Based on novel data, I have shown that the data usually put forward to deal with projection (e. g. "There are no kikes") are confounded by metalinguistic factors.

I thus considered another approach to slurring concepts, appealing to so-called "response-dependent" concepts. Response-dependent concepts - typically secondary quality concepts such as RED - have the interesting property to be inherently connected to non-conceptual, purely cognitive responses. Moreover, their extension is determined via the possessor's sensitivity to certain features of her environment. These crucial properties of response-dependent concepts make them excellent candidates for a model of slurring concepts.

Thus, I focused on the notion of response-dependence and developed the two important notions of *opacity* and *reflexivity*, so as to construct a response-dependent account of slurring concepts based on the model of RED.

I assessed the pros and cons of this account with regard to its ability to handle our (updated) list of explananda. I faced the need to add some complexity in the response involved in slurring concepts in order to explain the apparent categorization behavior of possessors which crucially differs from that of possessors of RED. Indeed, possessors of RED rely on their perceptual response to categorize an object as red, whereas it is unlikely that racists similarly rely on their racist response to categorize their targets as members of the target group.

I have then explored the possibility of giving to the notion of stereotype - as understood in statistical terms - a role in the response itself, but we saw that this was not fully satisfactory because categorization does not seem to rely on statistics. This discussion has led us to consider another response-dependent account giving away the property of reflexivity.

After a discussion on reflexivity and non-reflexivity, I have developed such a non-reflexive response-dependent account, based on the model of POLITE. Under this view, it is possible to possess a slurring concept even in the absence of the right sort of cognitive response. Although it addresses some of the problems raised against the first response-dependent account, we saw that such an account lost the initial interest we had to appeal to the notion of response-dependence, which was the inherent link between the concept and the non-conceptual response.

I then developed a potentially more satisfactory account of slurring concepts as essentialist concepts. Under successively two understandings of the notion of essence - one modal and another Aristotelian -, I have put forward the view that slurring concepts postulate an essence in their targets, and that this essence is taken to have a negative value. The combination of these two theses - the "Essentialist account" - has the resources to account for most, if not all, of the explananda we started with, I have argued. Because human essence is positive, and because SCs ascribe negative essences to their targets, SCs are dehumanizing.

Appendix

Appendix to 3.3: Three Remarks on Nunberg's Account

Circularity?

An additional potential problem with Nunberg's account relies on the fact that racists do not systematically use STs when talking about their targets. They do so only when they engage in a certain type of (racist) discourse. When talking with out-group members for instance, they might monitor their language, or might occasionally use the "negotiated default" even in conversation with in-group members.

But we saw that a crucial step in a Nunberg-like account is the recognition that a certain term "belongs" to a certain group. It is indeed the choice of such a term rather than that of the negotiated default which triggers the implicature to the effect that the speaker affiliates with the white racists, shares their attitudes toward black people and so on.

But there is an issue here: how can we identify the group that the ST belongs to, if members of the group use the term only in racist discourse? Is there a way to identify racist discourse independent of the use of such words? If not, the account would be circular.

And even more: if what makes "nigger" derogatory is the fact that racists use them in racist discourse, or to assert their white identity, then why do racists use this word in the first place? It cannot be that they use this word because of its derogatory force, as it has its derogatory force precisely because they use it[143].

If "nigger" has an ordinary descriptive semantics, it seems that racists had no reason to prefer this term to another term, or at least an additional story about the generation of these terms needs to be told. The link that Nunberg establishes between racist attitudes and racist words relies fully on one's knowledge that the very people who use the word are the bearer of the attitudes.

[143] I owe this observation to F. Recanati.

If Nunberg's deflationary approach aims at evacuating all potential pejorative conventional content from STs, it is not clear how a Nunberg-like approach could refine the link between racist attitudes and the content of a word without thereby incorporating attitudinal content in the word. There are many possible answers to this worry about a risk of circularity, I am here simply pointing at a potential need for elaboration.

Identifying the Relevant Group

There is another problem related to the question of the identification of the relevant group. For the derogatory term to trigger the relevant sort of implicature that Nunberg describes, it must be linked to the group to whom it belongs at some point in the derivation.

To this effect, the relevant group must be a sufficiently distinct, salient and recognizable social entity. Nunberg even goes as far as requiring that these groups are *self-counscious*, thus intending to explain why there aren't STs for dogs:

> dog haters don't constitute the kind of self-conscious collectivity who are going to come up with their own distinctive name for dogs (Nunberg 2016, p. 42)

But why should the relevant community that the racist is affiliating herself to be the sort of self-conscious social group he describes? Doesn't this wrongly predict that a term is a ST only when there is a self-conscious social group sharing attitudes towards the targets? Is that really the case of most STs? Do anti-Semitic people form a self-conscious social group? Do misogynistic males form a self-conscious social group?

And what about the term "cur" then, which refers to dogs and seems to express contempt towards them? Isn't that a ST for dogs? Maybe it isn't, but again, what is the relevance of an account which explains why "kike" is derogatory, but neither "cur" nor "bitch" nor "dirty jew"? It seems that Nunberg's account under-generates[144].

[144] I do not mention here the complications that the phenomenon of appropriation could bring to a Nunberg-like account, because there are ways to accommodate this aspect into a

Here are three examples showing that Nunberg's account also over-generates[145]. First, take the German term "Führer", which was used to refer to Adolf Hitler by a distinctive and self-conscious social group: the Nazis.

Given Nunberg's mechanism, an utterance of the term "Führer", as opposed for example to "Adolf Hitler", would trigger an implicature to the effect that the utterer affiliates herself with the Nazis. As Nazis are racists, that makes "Führer" a ST, on a par with "kike" or "nigger".

Similarly, back in the time, the term "Aryan" had a reading under which it roughly meant something along the lines of "non-Jewish", with a positive undertone. Was "Aryan" any one who was not Jewish, and that was considered a good thing. "Aryan" is thus a positive slur targeting non-Jewish people. But as "Aryan" is clearly a piece of jargon belonging to the Nazis, and as Nazis are anti-Semitic, "Aryan" is also an anti-Semitic and a racist ST.

Under Nunberg's view, "Aryan" is both a positive and a negative slur, with different targets. Third, going back to the example of "chocolatine", if it happened to be the case that all Southwest French people were anti-Semitic, "chocolatine" would be as much an anti-Semitic ST as "youpin", the French equivalent of "kike". These are unwanted consequences. Nunberg accounts overgenerates, because "Führer" is not a ST, "Aryan" is not both a positive and a negative racial ST, and it doesn't look like "chocolatine" could be a racial ST[146].

deflationary story. After appropriation of a ST by its targets, even though it is no longer solely a group of racists who use the term, it could still "belong" to the racists, or there could be a competition, etc.

[145] Thanks to B. Spector for his insight on that topic (p.c.).

[146] Note that it is not that clear that "chocolatine" could not be a ST. That intuition might stem from a failure to imagine Southwestern France as vastly and notoriously anti-Semitic. But imagine there is a word for bagels, or for cheesecakes, that only neo-Nazis use. Then it becomes easier to imagine that term as an anti-Semitic slur. I did not find a real life example of that, but would welcome any.

This is an attempt at providing a counter-example to Nunberg's view that the expressivity of STs is reducible to the recognition of a (primary) intention to affiliate with a racist group. Consider (172):

(172) John is Jewish, and I am anti-Semitic/I hate Jewish people/I belong to the Nazis.

Does'nt the explicitation of one's belonging to a group constitute an affiliatory speech-act as well? If yes, why is (173) way more powerful and expressive than (172)? This should be puzzling for Nunberg's account, I argue, because it predicts that "I am anti-Semitic" should be as expressive as "kike", since both constitute an affiliatory speech act.

(173) !John is a kike.

Consider also the term "non-Aryan" as applying to Jewish people (see above). Is an utterance of "John is non-Aryan" as expressive as (173)? It surely is an anti-Semitic utterance, but is that the sort of expression of subjective attitudes one is after?

In previous sections, we introduced T-terms and compared them to S-terms. There might be other potential contrastive linguistic features between slurring and thick concepts, further separating the two. Start with complex embeddings such as disjunctive (174)-(175) and conditional (176)-(177) presupposition filters, as well as certain attitude verbs (178)-(181):

(174) a. !Either Jewish people are not despicable, or this person is a kike.

b. Either there is nothing wrong with sexually explicitness, or this movie is lewd.

(175) a. !Either I am not anti-semitic, or this person is a kike.

b. Either I don't find sexually explicit behavior particularly wrong, or this movie is lewd.

(176) a. !If Jewish people were despicable, then this person would be a kike.

b. If there was something wrong with sexually explicit behavior, then this movie would be lewd.

(177) a. !If I were anti-semitic, then this person would be a kike.

b. Either I found sexually explicit behavior wrong, then this movie is lewd.

Considering only these two candidates for paraphrasing the evaluative content, one observes here a first contrast between STs and thick terms. It seems that the anti-Semitic content of "kike" takes wide-scope in both disjunctive and conditional filters, whereas it is way less clear for the evaluation associated with "lewd". These introspective judgments do not seem very clear cut, it would be worth testing it experimentally.

Another contrast can be observed with attitude verbs. Comparing the projection behaviors of thick terms and STs, Väyrynen also remarks that

> one difference concerns belief reports. We seem to find it acceptable to utter such reports as 'The Pope believes that the Rio carnival is lewd' even if we find lewd as used by the Pope objectionable. Reports of analogous utterances involving ethnic

> slurs tend to be found much less acceptable, increasingly so as the slur in question becomes more explosive. (Vayrynen, 2012, p. 11, footnote 22)

Here are four other examples:

(178) a. !Mary believes that Paul realizes that kikes are tall.

b. Mary believes that Paul realizes that lewd movies are forbidden.

(179) a. !In the fifties, black people were considered to be niggers.

b. In the fifties, *A streetcar named desire* was considered to be lewd.

(180) a. !Max doubts that Chomsky is a kike.

b. Max doubts that *Lolita* is lewd.

(181) a. !Hitler believed that Einstein was a kike.

b. The pope Pius VII believed that Sade's writings were lewd.

Here again, it seems that (178a)-(181a) are racist utterances, whereas (178b)-(181b) do not to commit the speaker to the (prude) attitudes associated with lewd. The following will help us understand what is at stake in these contrasts, to which we will come back at the end of the discussion.

Again, there seems to be a contrast, which could well deserve experimental validation as well. Also, the extent to which the contrast is due to the taboo component of slurs is unclear. Many slurs seem in fact to have acquired a high degree of toxicity (Anderson and Lepore, 2013a, 2013b), so that any utterance containing their phonological form might break transgress sort of a norm and hence trigger additional pragmatic inferences.

Hay 2011, focusing on the respective behavior of STs and of what he calls "general pejoratives" (e. g. "jerk", "asshole") under attitude reports, reaches similar conclusions in noticing that

> general pejorative terms - like "jerk" - have descriptive components that are not detachable, and, when embedded in belief reporting sentences, the negative attitudes they are used to express get attributed to the subjects of such sentences. In contras, slurs […] have detachable descriptive components, and they can be used by speakers

> to express *their own* negative attitudes even when reporting the beliefs of others. (Hay 2011, p. 21)

Now, the discussion on objectionable vs. non-objectionable thick concepts might also be used as a distinctive feature of the thick, as opposed to STs, because the negative evaluation associated with STs seems to be always objectionable.

But there is in fact variation, as the racists do not find the evaluation in STs objectionable. The difference is one of degree rather than of kind, and the warranted-unwarranted is just dimension that thick concepts explore more than STs. There are two another such dimension that thick concepts seem to vary along more than STs do: that of the *polarity* of the evaluation, and that of the *intensity* of the evaluation.

Many have remarked that there is no such thing as a "positive" ST. That is, a ST picks out a referent and conveys an evaluation about it that is always negative. That is indeed the case of all STs that are discussed in the literature, and of all STs we discussed so far.

On the opposite, we find thick terms on both poles of the evaluation spectrum. There are many negative thick terms, like "lewd", "nasty", or "cruel", and many positive thick terms, like "courageous", "chaste", or "kind". Why would we not find cases of positive STs?

First of all, I shall point at that, although it is true that it is hard to find cases of positive STs (and that seems to be the case of most languages), it is not clear that there are none. It is not clear that a term like "Aryan", as used in Nazi Germany, was not tinted with positive evaluations, picking out non-Jewish people under one reading, or imaginary supermen on another. Or in modern European French, the word "savant" can be used for (a subclass of) scientists, and seems to be positively loaded too. "Saint" might arguably be another such positive ST.

But let us assume that, even if there might be an occasional positive ST, there is at least a huge imbalance between positive and negative STs, contrary to thick terms. Now, I shall say that that difference cannot tell us much about a potential structural difference between STs and thick terms, because it could well be that this contrast normally follows from the difference in these terms' descriptive parts.

By agreement, we tend to call "slurs" only these hybrid evaluative terms that target groups and individuals as social entities. On the contrary, thick terms refer to a way broader and more diverse set of entities, from behaviors to actions and so on.

Given this difference, many non-linguistic, social and evolutionary factors might intervene in explaining why there aren't positive STs. Maybe it is a fact of one's social cognition that it's more rare to have positive feelings towards an identified social groups than to have negative feelings.

There are many unresolved possibilities of that sort which could explain, on a non-linguistic basis, why there are fewer positive STs than positive thick terms. As soon as there is such a specific social target for STs, and not for thick terms, the difference in proportion of negative and positive instances becomes linguistically irrelevant.

A last putative contrast between STs and thick terms has to do with *gradability*. Gradable adjectives like "cold" or "frightened" can be combined with expressions like "very" or "a bit", and can be put in comparatives like "more x than" or "as x as". One can be "a bit frightened", "very cold" or "taller than Mary", but not "very married", "a bit wooden" or "less dead than Mary".

Similarly, most thick terms seem gradable, whereas most STs seem non-gradable, as the contrast between (182) and (183) suggests:

(182) a. That movie is very lewd.

b. That movie is a bit lewd.

c. That movie is more lewd than the previous one.

(183) a. *!This person is very kike.

b. *! This person is a bit kike.

c. *!This person is more (of a) kike than John.

An obvious response would be that STs are usually nouns, and that nouns aren't gradable in the same way adjectives are. The present contrast in gradability would amount to a difference in syntactic category, which is nothing of an enlightening structural distinction between the two kinds of evaluatives.

Some NPs like "stamp-collector" or "idiot" are said to be gradable, because they can undergo a degree modification by expressions like "big", or "enormous" for instance (Morzycki, 2009). One can be an "enormous idiot" or a "big stamp-collector", not an "enormous German" or a "big man" (at least not in the expected sense that ones degree of Germanness or manness is high).

But although they are gradable, such nouns can of course not be grammatically combined with "very", "a bit" or "more x than" constructions. Take the thick term "coward" for instance, which is gradable as nouns are, not as adjectives are:

(184) a. *This person is very/a bit coward.

b. John is an enormous/big coward.

Gradability amounts to sensitivity to degree modification, rather than to licensing in specific linguistic environments. So, is it the case that STs are not sensitive to degree modification? Consider:

(185) a. !John is a huge faggot/dyke.

b. *!John is an enormous/big nigger/chink.

c. ?!John is an enormous/big kike.

The data seems to be heterogeneous. It could be that when the target is identified through her nationality or ethnicity, which are categorical properties, then the ST is less gradable than when the target is identified through her behavior or other non-categorical properties. Be it or not, at least one sees that some STs are gradable as nouns, and that the alleged contrast in gradability is here again of degree than of category.

To sum up, there are three dimensions that thick terms seem to spread along more than STs do: a *warranted-unwarranted* dimension, a *positive-negative* dimension and a *weak-strong* dimension.

These differences seem to be differences of degree rather than of kind, but they reinforce the distinction I introduced in chapter 5 between thick evaluative terms and concepts on the one hand, and STs and SCs on the other hand.

The two seem to be hybrid evaluatives, and might be linked diachronically, but their relation to their "counterpart" is crucially different. Thick terms and concepts seem to target only a subclass with bad properties, but STs and SCs target a whole class.

It might seem that another interesting advantage of an opaque response-dependent account of SCs such as RDA is that it could be a way to flesh out the notion of perspective that Camp (Camp 2013) argues plays an essential role in slurring representations:

> I want to argue that slurs are so rhetorically powerful because they signal allegiance to a perspective: an integrated, intuitive way of cognizing members of the targeted group.

What is exactly this "integrated, intuitive way of cognizing members of the targeted group"? She goes on:

> A perspective is representational, insofar as it provides a lens for interpreting and explaining truth-conditional contents, but it need not involve a commitment to any specific content. Likewise, a perspective typically motivates certain feelings as appropriate to feel toward its subject, but it is not itself a feeling. In a general sense, then, my suggestion is that slurs are akin to other expressions part of whose conventional function is not merely to refer or predicate, but to signal the speaker's social, psychological, and/or emotional relation to that semantic value. (Camp 2013, p. 335)

She adds that "the notion of perspective is fairly intuitive but rarely spelled out". Our conception of opaque response-dependence can be seen as a way to spell out the notion of perspective. Indeed, according to RDAred, BOCHE and GERMAN could happen to be coextensive, but they come with very different canonical modes of presentation. What distinguishes the two concepts is therefore their mode of presentation of the referent: one involves an emotional response, the other does not.

If opaque response-dependent concepts involve perspectives, they should be sensitive to perspectival effects in thought and language, such as free indirect speech for instance. This is what I will evaluate now.

Let us then take a clear opaque response-dependent concept, like the coarse and vulgar French BONNASSE. "Bonnasse" is a heavily sexist term used only by certain young, crude male (and maybe, to some extent, some women) chauvinists. They apply it to females whose physical appearance arouse them. Users of this term would typically share utterances like "This chick is such a *bonnasse*", emphasizing the sexual attraction they feel for the targeted women[147].

It is safe to consider this term a good example of a term expressing an opaque response-dependent concept. Few people use the term, and those who do surely possess the associated concept. BONNASSE applies to individuals (certain females) whose specific anatomical properties trigger in some relevantly equipped cognitive systems (mostly certain obnoxious sexist males) a specific sort of (sexual) response.

It is clear that normal users of the term are the responders, so that when a clear non-responder (like a women in the relevant context) uses the term speakers know it is a special, perspectival or echoic use. Since the term is response dependent, it involves a particular perspective whenever it is used, that is, the perspective of the responders[148].

Consider now the behavior of "bonnasse" in free indirect discourse. Imagine Jean is a sexist male, and the narrator writes:

(186) !Jean ouvrit la fenêtre. Où cette bonasse était-elle passée ?

Jean opened the window. Where this hot-chick was-she passed

Jean opened the window. Where could this hot-chick be now?

In that case of free indirect discourse, the narrator is reporting Jean's train of thought,[149] and the term "bonasse" is evaluated through Jean's mind in some sense: the associations are not attributed to the narrator. The narrator of (186) could for example very well be a woman. In

[147] The closest English equivalent of the French "bonasse" I can think of is something along the lines of "hot chick".

[148] See Camp 2013 for an account of slurs in terms of perspectives. RDA could be seen as an attempt to flesh-out Camp's notion of a perspective in that debate.

[149] See Eckardt (2015) for an extensive study of free indirect discourse, its markers and semantic interpretation.

contrast, someone uttering only the second part of (186) would surely not be a woman, at least if the utterance is not echoic or ironic in some sense.

Now, there are other possible perspectival effects that are not directly reducible to free indirect discourse, in cases where it is not someone's actual thoughts that are displayed, but thoughts she might have had. Consider for example (187):

(187) !Jean ouvrit la fenêtre. Il ne savait pas qu'une bonasse venait de passer.

Jean opened the window. He NEG did-know NEG that a hot-chick came to pass

Jean opened the window. He didn't know that a bonasse just passed by.

Since in (187), one is talking about something Jean ignores, one is not reporting or displaying any of Jean's actual thoughts. Nonetheless, it seems that the attitudes associated with use of the term "bonnasse" is not attributed to the narrator, but to Jean again.

Might Jean come to be acquainted with the women one is talking about, he would have the relevant response, and call her, or think of her, as a "bonnasse". In free indirect speech, as well as in other perspectival effects like (187) - call such uses "potential free indirect speech", uses of response-dependent terms are not ascribed to the narrator, that is, they don't scope out of these sort of metalinguistic operations.

We saw earlier that STs seem to scope out of all truth-conditional and intensional operators, and that only metalinguistic operators seemed to be able to capture their effects under their scope.

But the behavior of STs contrast with that of "bonnasse", as we see in (188) and (189) testing the projection behavior of STs in free indirect discourse and potential free indirect discourse, in parallel with (186) and (187):

(188) !Hitler opened the window. Where could this kike be now?

(189) !Hitler opened the window. He didn't know that a kike just passed by.

Unlike in (186) and (187), where the attitude associated with "bonasse" was associated to the character rather than to the narrator, it seems that the anti-Semitic attitude expressed by "kike" in (188) and (189) still projects to the higher level: it seems that only an anti-Semitic narrator could have couched such a sentence.

This contrast suggests that, after all, SCs might not be perspectival in the required sense. This might count as an additional explanatory cons of RDAred.

A Collection of Slurring Terms and Other Insults

Here is a non-exhaustive list of STs and similar terms ranked by properties of their respective targets.

- Ethnicity and race

Beaner, boche, camel jockey, canuck, cheesehead, chinaman, chink, coconut, coon, cracker, curry-muncher, dago, frog, gook, gorilla, goy, greaser, gringo, half-breed, haole, hebe, heeb, hillbilly, honky, hymie, injun, jap, jigaboo, kike, kimchi, kraut, limey, macaca, mick, negro, nigger, nip, paki, pickaninny, polack, potato head, redneck, russki, sand nigger, shiksa, slant-eye, spade, spaghetti-eater, spic, towel-head, wetback, wog, wop, yankee, yellow, yid

- Sex, sexual orientation, gender, marital status

Boy, breeder, broad, bugger, carpet-muncher, chick, cunt, dame, dyke, fag, faggot, fairy, floozy, fruit, girl, lecher, lesbo, lothario, pansy, perv, queen, queer, rake, rice queen, sod, twink, sissy, slit, slut, skirt, spinster, tart, tramp, wench, wuss

- Morals

Bimbo, floozy, harlot, prig, prude, slut, strumpet, whore

- Politics

Commie, fascist, facho, gun nut, leftie, Nazi, peacenik, pinko, radical, reactionary, right winger, tea bagger, tree hugger

- Religion

Bible-thumper, christ-killer, clamhead, firewood, heathen, heretic, holy roller, infidel, Jesus freak, Jewish American Princess (JAP), kike, lamp shade, mackerel snapper, papist, raghead

- Health, age, appearance

Bean pole, blimp, crip, dwarf, four-eyes, geezer, gimp, hag, handicapped, hippo, lardass, midget, pig, punk, shrimp, slob, string bean

- Substance abuse

Acid freak, boozer, crack head, dope fiend, druggie, freak, hophead, junkie, lush, meth head, wino

- Popularity

Nerd, geek, dweeb, loser

- Occupation, profession, financial

Bean counter, bum, charlatan, con artist, cotton picker, crook, deadbeat, demagogue, drudge, empty suit, flunky, fuzz, hack, hood, jungle bunny, gigolo, gold-digger, goon, grifter, hatchet man, ho, hooker, huckster, hustler, kunta kinte, lamp shade, leech, loan shark, money-grubber, narc, oven, paper shuffler, pencil pusher, peon, pig, pimp, prole, quack, scab, scrub, shrink, shylock, shyster, skinflint, snitch, spendthrift, sponge, stool-pigeon, suit, thug, tightwad, whore

- Life-style, character

ass-kisser, boor, brown-noser, bum, chicken, couch potato, dork yahoo, dweeb, flake, freak, freeloader, fuddy-duddy, geek hippie, hick, jerk off, kook, lame ass, mouse, nerd, old fogy, party pooper, patsy, quitter, riffraff, rube, redneck, slacker, square, staggler, stick in the mud, tight-ass, toady, trailer trash, twerp yuppie, weenie, weirdo, wiener, wimp, wuss, yokel

- Intelligence/sanity

Airhead, bimbo, birdbrain, bonkers, bozo, buffon, chowderhead, clodhopper, crackpot, cretin, dabbler, deranged, dolt, dope, dickhead, dilletante, dingbat, doofus, dumbass, dumbfuck, dunce, dupe, egghead, egomaniac, fool, idiot, ignoramus, imbecile, lunatic, maniac, meathead, moron, nincompoop, nitwit, numskull, nut, nut case, nut job, patsy, pervert, philistine, pigeon, psycho, psychotic, retard, rube, sap, sociopath, sucker, twit, wacko

- Fictional entities

Fangs, furface, moondog, pointy-ear, toaster

- Vulgar, annoying, inconsiderate people

Asshole, bastard, bitch, blowhard, brat, creep, cunt, dick, dirtbag, douchebag, fart, fink, jerk, kvetch, loudmouth, louse, nag, pain in the ass, pest, prick, punk, rat, rat fink, schmuck, scumbag, scuzzball, shit, shithead, sleezeball, slimeball, smart-ass, snake, snitch, snot, SOB, stool, swine, twerp, twit, windbag

- Proper names, animals, others

Ape, barbarian, beast, benedict Arnold, brute, cow, dog, four-eyes, guido, Hitler, hog, hymie, ikey, inyenzi, Jezebel, Judas, judenschwein, La la land, martinet, mick, Neanderthal, pig, porch monkey, rat, Quisling, ranga, savage, sheboon, snake, Stalin, stinkpotter, toad, troglodyte, swine, worm

- Related Terms

- *Epressive Intensifiers:* blessed, blasted, darn, damn, goddam, effin', freakin', fuckin', motherfuckin'

- *Exclamations:* Shit! Dammit! Fuck! Goddam it! Oh crap! Holy shit!

- *Laudatives:* angel, saint, sweetheart, babe, hottie, knockout, hunk, artist, pro, ace, whiz

For more, the reader can look at the Racial Slur Database.

Glossary

Absurd Counterfactual Conditional (ACC): A special use of conditionals, where the patent falsity of the consequent implicates the falsity of the antecedent (e. g. "If homeopahty cured Mary, then I am the queen of England").

Accidental surface properties: Accidental surface properties are observable and not connected to the essential property in any way. They do not play a role in categorization. The fact that cats are made out of a lot of water (like all living things) constitutes an accidental surface property for instance.

Asymmetric Dependence Heuristic (ADH): Kinds of ST-uses/SC-deployments that plausibly stand in an asymmetric existence-dependence relation to other ST-uses/SC-deployments are (likely to be) parasitic.

Canonical Mode of Presentation: a concept's canonical mode of presentation is the Normal mechanism it is grounded on and which provides its Normal possession conditions.

Central Cases Definition (CCD): Central cases of slurring concepts are the cases in which they accomplish their proper function.

Co-Description Thesis (CDT): Evaluative concepts have the same reference-fixing, descriptive content as that of their neutral counterparts.

Co-Extensionality Thesis (CET): STs have the same extension as their NCs.

Conceptual Hypothesis (CH): STs Normally express SCs.

Conventionality thesis (CT): The evaluative component of evaluative concepts is just conventionally (hence arbitrarily) associated to their reference-fixing, descriptive component.

Default Maxim: Use default vocabulary unless there is a reason not to.

Derivative essential properties: Derivative essential properties are not directly derived from the essence, but they accompany essential properties. They roughly correspond to what evolutionary biologists call "spandrel". For instance, the eye's blind spot is a consequence of human DNA (hence of the essence), but it was not really selected for its advantages in fitness, it is rather a necessary bad accompanying the eye. The cat's blind spot thus belongs to its derivative essential property, whereas its fur or claws are direct essential properties.

Essences: An essence is whatever is responsible for *i)* kind membership and *ii)* an array of observable features (be it metaphysically real or not). The essence of a cat could be it's DNA for instance. There is room for physically instantiated essences such as DNA, but as we shall see later, also for historical or social essences.

Essence1: Something is an essence iff it is a property that an object necessarily possesses at all worlds.

Essence2: An essence is the nature of a thing or kind of thing. The essential properties of a thing are those that derive (in the right way) from its essence. Crucially, a thing can exist without manifesting its essential properties.

Essentialism: Essentialism is the view that some objects have essential properties. Psychological essentialism is the view that human naive psychology is basically essentialist.

Essential properties: Essential properties are (maybe causally) derived from the essence. Subjects tend to rely on the detection of essential properties to infer the presence of an underlying essence. The cat's claws or its fur could be such essential properties for instance. The essential properties associated with SCs correspond to the stereotype.

Essentialist Account (EA): *i)* SCs are essentialist concepts; *ii)* SCs attribute negative essences.

EAaristotle: i) SCs attribute an underlying normal cause to the syndrome of their target's; ii) This cause is an essence and has a negative valence.

EAputnam: i) SCs attribute necessary properties to their targets; ii) The necessary properties that SCs attribute to their targets are negative.

Generalized Reflexivity Thesis: possession of an empirical concept F requires having the associated canonical response R.

Hybrid Expressivist Accounts (HEA): Hybrid expressivist accounts of STs subscribe to the CET and call on other dimensions of meaning to account for their additional expressive properties.

Immunity Definition of Projection (IDP): "An implication projects iff it survives as an utterance implication when the expression that triggers the implication occurs under the syntactic scope of an entailment-cancelling operator." (Beaver and Roberts 2010)

Moral claim: necessarily, no one ought to be the target of negative moral evaluation because of belonging to a group.

NC Constraint (NCC): Slurring representations only emerge when they have NCs. There are actually four distinct subversions of this constraint:

i) STs require NC-Ts

ii) STs require NC-Cs

iii) SCs require NC-Ts

iv) SCs require NC-Cs

Neutral Counterparts (NCs): A representation is a Neutral Counterpart (NC) of a hybrid representation when it shares its descriptive component and lacks its attitudinal component. For instance, JEW is the NC of the SC KIKE.

Neutral Counterparts' (NC's): A representation is a Neutral Counterpart (NC) of a hybrid representation when it shares its target and lacks its attitudinal component.

Normal Conditions: The normal conditions of emergence of an item are, roughly, the external (E) and/or internal (I) conditions under which the fulfilment of their proper function allows for its reproductive success. In the case of SCs, I consider two sorts of conditions: E-conditions i. e. social conditions; I-conditions i. e. psychological conditions.

Normal possession: The subject's possession of a concept is *Normal* when it relies on the conditions under which the fulfillment of its proper function allowed for its reproductive success[150].

Opacity: A response-dependent concept is *opaque* when knowledge of the governing RDB is not necessary for subjects to possess the concept. The RDB is thus not *a priori*.

Possession-Response rule for SCs (PR): Possessors of a SC are also responders.

Possession-Response rule for SCs (PR): Possessors of a SC are also responders.

Projection (Immunity Definition, IDP): "An implication projects iff it survives as an utterance implication when the expression that triggers the implication occurs under the syntactic scope of an entailment-cancelling operator." (Simons et al. 2010)

Projective-expressive content: All the things that are insensitive to truth-conditional operators because they are expressive/non-conceptual.

Projective-filtering content: All the things that are insensitive to truth-conditional operators because they are not at all in the conventionalized content of the utterance.

Projective-layering content: All the things that are insensitive to truth-conditional operators because, even though they are encoded in the truth-conditions, they are presented as not being the main point of the utterance.

Proper Function:

> A proper function of [...] an organ or behavior is, roughly, a function that its ancestors have performed that has helped account for proliferation of the genes responsible for it, hence helped account for its own existence. (Millikan 1989, p. 289)

A similar notion of proper function can be extended to representations. In the case of SCs, the proper function will be the function that is responsible for its creation and proliferation.

[150] The color-blind color-scientist's possession of the concept RED is thus not Normal. The Normal possession of RED is grounded on red*, not on scientific expertise.

Radical/Modest Conceptualism: Radical conceptualism maintains that an account of STs that does not take into account underlying psychological states is *incorrect*. Modest conceptualism maintains that an account of STs that does not take into account the underlying psychological states of speakers is *incomplete*.

Reference-Based Evaluation (RBE): (i) T-terms express concepts with a descriptive component *richer* than that of their counterparts; (ii) T-terms and concepts refer to "bad" *subgroups*; (iii) Their evaluative content is (extra-semantically) associated with the perceived negative properties of the subgroup. This view challenges both the co-description and the conventionality theses (CDT and CT) of hybrid expressivists theories (HEA) *à la* Frege.

Reflexivity: A response-dependent concept is *reflexive* when its possessors possess the concept in virtue of being responders.

Reflexivity Thesis (for RED): it is not possible to possess RED without having the response red*.

Response-Dependence (RD): A concept is response-dependent when it picks out a dispositional property of an object to elicit a mental response from an agent under specified conditions. RED is the most typical example, as it (arguably) applies to the power certain objects have to elicit a perception of red in healthy human beings under standard lighting.

RDA: SCs are Response-Dependent Concepts

RDAred: x is a *SC* iff x would cause NC-phobic people to have a yuk*(i)-qua(ii)-NC(iii) response under [normal conditions of slurring thought].

RDApolite: it is *a priori* that x is a SC iff x would cause NC-phobic people to have a worthy-of-yuk*(i)-*qua*(ii)-NC(iii) response under Normal conditions of slurring thought.

Semantic claim: The pejorative content of SCs is part of their truth-conditional content (e. g. WOP = PEJ(ITALIAN)).

Slurring Concepts (SCs): Private psychological slurring representations.

Slurring Concepts Hypothesis (SCH): Pejoratives are not merely a matter of speech. There are pejorative concepts in thought (and Slurring Terms express such concepts).

Slurring Deployments (SDs): Deployment of SCs, or of any concept that fulfills the function of an SC, in thought.

Slurring Terms (STs): Public linguistic slurring representations.

Slurring Uses (SUs): Uses of STs, or of any term that fulfills the function of an ST, in language.

Separability: A seemingly hybrid concept is separable when its descriptive and attitudinal components are independently characterizable. In particular, a characterization of a separable concept's extension need not invoke the concept's attitude.

Slurring Concepts Hypothesis (SCH): Slurring is not merely a matter of speech. There are slurring concepts in thought (and Slurring Terms express such concepts).

Surface Stereotype View (SSV): Slurring concepts are to be identified with negative stereotypes of their target categories.

Syntax/semantics Definition of Projection (SSDP): An inference projects iff it is semantically interpreted above the scope of an entailment-cancelling operator it is syntactically embedded under.

Target: The target of a slurring representation is the group or individual it is meant to apply to.

T-terms and T-concepts: terms/concepts targeting certain *subgroups* in virtue of (*i)* a reference to (at least) the *supergroup*, and (*ii)* an additional *descriptive* element[151] motivating (*iii)* an evaluation.

Truth-conditional-evaluative view (TCE): STs express concepts with a rich reference-fixing component, including a standard predicate and an evaluative operator.

[151] For comparison, recall the definition of STs (which is here not restricted to the specific subcase of slurs):

> **S-terms**: terms whose meaning is hybrid (it is made of at least two different kinds of meaning) and whose meaning components are separable (one can find or construct neutral counterparts)

Truth-Conditional-Stereotypical-Evaluative view (TCSE): STs express concepts with a rich reference-fixing component, including a stereotype-predicate and an evaluative operator.

Use-Approval rule (UA): Language users who are disposed to actually use STs are usually disposed to actively deploy in thought the SCs that the terms they use express, without reservations.

Use-Possession rule (UP): Language users usually possess the concepts that the terms they use express.

Use-Possession rule (UP'): Language users usually possess the SC that the ST they use express.

References

A

Anderson, L. & Lepore, E. (2013a), Slurring Words, *Noûs* 47, 25–48.

Anderson, L. & Lepore, E. (2013b), What Did You Call Me? Slurs as Prohibited Words, *Analytic Philosophy,* Vol. 54, No. 3, pp. 350–363.

Andreasen, R. O. (1998), A new perspective on the race debate. *The British journal for the philosophy of science*, *49* (2), 199-225.

Appiah, K. A. (1996), *Race, culture, identity: Misunderstood connections*, Tanner Lectures on Human Values, 17, 51-136.

Armstrong, D. M. (1969), *Color-realism and the Argument from Microscopes*, in Brown and Rollings 1969, pp. 119-31.

Austin, J. L. (ed. and trans.), (1974), *The Foundations of Arithmetic, A logic-mathematical enquiry into the concept of number*, Oxford: Blackwell, second revised edition (second edition, 1953; first edition, 1950).

Austin, J. L. (1975), *How to do things with words*, Oxford University Press.

Ayer, A. J. (1936), Language, Truth and Logic, London: Gollancz 2nd Edition, 1946.

B

Bach, K. (1999), The Myth of Conventional Implicature, *Linguistics and Philosophy*, 22(4): 327–366.

Barker, C. & Taranto, G. (2002), The paradox of asserting clarity, in *Proceedings of the western conference on linguistics* (WECOL), 14, 10-21.

Beaver, D. I. & Geurts, B. (2011), Presupposition, *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University.

Bechara, A., Damasio, H., and Damasio, A. R. (2000), Emotion, decision making and the orbitofrontal cortex, *Cerebral cortex*, 10(3):295–307.

Bell, M. (2013), *Hard Feelings: The moral psychology of contempt*, Oxford University Press.

Benveniste, E. (1958), *Les verbes délocutifs*, Mél. Spitzer, Berne, pp. 277-285.

Bianchi, C. (2014a), Slurs and Appropriation: an Echoic Account, *Journal of Pragmatics*, 66, 35-44.

Bianchi, C. (2014b), The speech-act account of derogatory epithets: some critical notes, in. *Liber Amicorum Pascal Engel*, pages 465–478.

Blackburn, S. (1984), *Spreading the word*, Oxford University Press Oxford.

Blackburn, S. (1985), *Errors and the phenomenology of value, Morality and Objectivity* [Routledge & Kegan Paul].

Blackburn, S. (1992), Through Thick and Thin, *Proceedings of the Aristotelian Society*, supplementary volume 66, 284-99.

Blackburn, S. (1993), *Essays in quasi-realism*, Oxford University Press.

Blackburn, S. (1998), *Ruling passions*, Clarendon Press Oxford.

Blum, L. (2002). *"I'm not a racist, but...": the moral quandary of race*, Cornell University Press.

Bolinger, R. J. (2015), The pragmatics of slurs, *Noûs,* 50(3).

Bott, L. and Noveck, I. A. (2004), Some utterances are underinformative: The onset and time course of scalar inferences, *Journal of memory and language*, 51(3), 437-457.

Boghossian, P. A. & Velleman, J. D. (1989), Colour as a secondary quality, *Mind*, 98(389), 81-103.

Brontsema, R. (2004), A queer revolution: reconceptualizing the debate over linguistic reclamation, Colorado Research in Linguistics, 17, 1-17.

Byrne, A. and Hilbert, D. R. (1997), Colors and reflectances, *Readings on Color* [MIT press], 1.

Byrne, A. and Hilbert, D. R. (2003), Color realism redux, *Behavioral and Brain Sciences*, 26(01):52–59.

C

Camp, E. (2011), Slurs, Semantics, and Stereotypes, Presented at the *Syracuse Philosophy Annual Workshop and Network 2011*, Syracuse University, Syracuse, NY.

Carnap, R. (1935), *Philosophy and logical syntax*, London: Kegan Paul, Trench, Trubner & Co.

Carroll, N. (1997), Simulation, Emotions, and Morality, *Beyond Aesthetics*: *Philosophical Essays*, Cambridge University Press.

Carston, R. (1996), Metalinguitstic negation and echoic use, *Journal of Pragmatics*, 25(3), 309-330.

Carston, R. (1997), Enrichment and loosening: complementary processes in deriving the proposition expressed?, In *Pragmatik* (pp. 103-127). VS Verlag für Sozialwissenschaften.

Chalmers, D. J. (1995), Absent qualia, fading qualia, dancing qualia, *Conscious experience*, pages 309–328.

Chemla, E. (2009a), Similarity: Towards a unified account of scalar implicatures, free choice permission and presupposition projection. Ms. LSCP & MIT.

Chemla, E. (2009b), Presuppositions of quantified sentences: experimental data, *Natural Language Semantics*, 17(4), 299-340.

Cepollaro, B. & Stojanovic, I. (2016), Hybrid Evaluatives: In Defense of a Presuppositional Account, *grazer philosophische studien*, *93*(3), 458-488.

Chierchia G. & McConnell-Ginet, S. (2000), *Meaning and Grammar: An introduction to Semantics*, Second Edition, Cambridge, MA: MIT Press.

Clark, H. H. (1992), *Arenas of language use,* University of Chicago Press.

de Condillac, E. B. (2001), *Essay on the origin of human knowledge*, Cambridge: Cambridge University Press.

Croom, A. (2011), "Slurs", *Language Sciences*, 33, 343-358.

Croom A. (2014), Spanish slurs and stereotypes for Mexican-Americans in the USA: A context-sensitive account of derogation and appropriation, *Sociocultural Pragmatics*, 2, 1-35.

Culpeper, J. (2011), *Impoliteness: Using Language to Cause Offense*, Cambridge: Cambridge University Press.

D

Darwall, S. L. (1977), Two kinds of respect, *Ethics*, 88(1), 36-49.

Davis, C. and McCready, E. (2016), Expressives in Questions, *Proceedings of SALT 26*, 000-000, 2016.

Dennett, D. C. (1991), Real Patterns, *The Journal of Philosophy*, 88(1), 27-51.

Dennett, D. C. (1993), *Consciousness explained*, Penguin UK.

DiFranco, R. (2015), Do Racists Speak Truly? On the Truth-Conditional Content of Slurs, *Thought: A Journal of Philosphy*, 4(1), 28-37.

Donnellan, K. S. (1966), Reference and definite descriptions, *The Philosophical Review*, 281-304.

Ducrot, O. & Vogt, C. (1979), De magis à mais : une hypothèse sémantique, *Revue de Linguistique Romane*, Lyon, 43(171-172):317–341.

Dummett, M. (1973), *Frege: Philosophy of language, The Interpretation of Frege's Philosophy*, Duckworth, London.

E

Ebert, C., Evert, S., Wilmes, K. (2011), Focus Marking via Gestures, In Reich, Ingo et al. (eds.), *Proceedings of Sinn und Bedeutung 15*, Saarland University Press: Saarbrücken, Germany.

Eckardt, R. (2014), The Semantics of Free Indirect Discourse: How Texts Allow Us to Mind-read and Eavesdrop, *Current Research in the Semantics/Pragmatics Interface*, 31, Brill.

Egré, P. & Magri, G. (2008), Presuppositions and Implicatures, *Proceedings of the MIT-Paris Workshop*.

Eklund, M. (2011), What are Thick Concepts?, *Canadian Journal of Philosophy*, 41(1), 25-49.

Evans, G. (1982), *Varieties of Reference*, Oxford: Oxford University Press.

F

Finlay, S. (2005), Value and Implicature, *Philosopher's Imprint*, 5(4): 1-20 (2005).

Fodor, J., Beaver, T., Garrett, M. F. (1974), *The psychology of language, an introduction to psycholinguistics and generative grammar,* New York: McGraw-Hill.

Fodor, J. (1981), Methodological Solipsism Considered as a Research Strategy in Cognitive Science, In Fodor, J. *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, Cambridge MA: MIT Press. 225-256.

Fodor, J. (1982), Cognitive Science and the Twin Earth Problem, *Notre Dame Journal of Formal Logic* 23, 97-115.

Fodor, J. A. (1998a), There are no recognitional concepts: not even RED, *Philosophical issues*, 9, 1-14.

Fodor, J. A. (1998), *Concepts: Where cognitive science went wrong*, Oxford University Press.

Foot, P. (1958a), Moral arguments, *Mind* 67: 502-13. Reprinted in Virtues and Vices, Oxford: Blackwell (1978).

Foot, P. (1958b), Moral beliefs, *Proceedings of the Aristotelian Society,* 59: 83-104.

Foot, P. (2003), *Natural Goodness*, Clarendon Press.

Fox, D. (2007), Free choice disjunction and the theory of scalar implicatures, In *Presupposition and implicature in compositional semantics*, ed. Uli Sauerland and Penka Stateva, 71–120. Palgrave-Macmillan.

Fox, D. (2014), Cancelling the Maxim of Quantity: Another challenge for a Gricean theory of scalar implicatures, *Semantics and Pragmatics* 7:1–20.

Frege, G. (1879), *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: L. Niebert; reprinted in Frege 1998a; trans. as "Begriffsschrift, a Formula Language, Modeled upon that of Arithmetic, for Pure Thought" in From Frege to Gödel, edited by J. van Heijenoort, Cambridge, MA: Harvard University Press 1967, and as "Conceptual Notation" in Frege 1972; selections in Frege 1997.

Frege, G. (1879-1891), "*Logik*", in Frege 1983: 1-8; translated as "Logic" in Frege 1979: 1-8.

Frege, G. (1892), Ueber Sinn und Bedeutung, *Zeitschrift fur Philosphie und Philosophische Kritik* 100, pp. 25-50. Translated as "On Sense and Reference" in P. T. Geach and M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*, Third edition. Oxford: Blackwell, 1980, pp. 56-78.

Frege, G. (1884), *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: W. Koebner, 1884. Complete translation by J. L. Austin in Austin (1964).

Frege, G. (1918), Der Gedanke, Eine Logische Untersuchung, *Beiträge zur Philosophie des deutschen Idealismus*, I:58–77. Translated as "Thoughts" in B. McGuinness (ed.), *Collected Papers on Math, Logic, and Philosophy,* Oxford: Blackwell, 1984. Reprinted in N. Salmon and S. Soames (eds.), *Propositions and Attitudes*, Oxford: Oxford University Press (1988), pp. 33-55.

Frege, G. (1979), *Posthumous Writings*, translated by Long and R. White, Oxford: Basil Blackwell.

G

Geach, P. T. (1965), Assertion, *The Philosophical Review*, pages 449–465.

Gert, J. (2006), A realistic colour realism, *Australasian Journal of Philosophy*, 84(4):565–589.

Geurts, B. (1996), Local satisfaction guaranteed: A presupposition theory and its problems, *Linguistics and Philosophy*, 19(3), 259-294.

Geurts, B. (1998), The mechanisms of denial, *Language*, 74, 274-307.

Geurts, B. (1999), *Presuppositions and pronouns*, Elsevier.

Gibbard, A. (1992), Thick Concepts and Warrant for Feelings, *Supplement to the Proceedings of the Aristotelian Society* 66: 267-83.

Gibbard, A. (2003), *Thinking How to Live*, Cambridge, MA: Harvard University Press.

Glasgow, J. (2010). *A theory of race*. Routledge.

Greenberg, J. & Pyszczynski, T. (1985), Compensatory self-inflation: A response to the threat to self-regard of public failure, *Journal of Personality and Social Psychology*, 49, 273-280.

Grice, H. P. (1957), Meaning, *The philosophical review*, 377-388.

Grice, H. P. (1975), *Logic and conversation*, in P. Cole & J. Morgan (ed.), Syntax and Semantic, 3 : Speech Acts, pp. 41-58, New York: Academic Press. Reprinted in H. P. Grice (ed.), Studies in the Way of Words, pp. 22-40, Cambridge, MA: Harvard University Press (1989).

Grim, P. (1981), *A note on the ethics of theories of truth*, In Vetterling-Braggin 1981: 290-298.

Gutzmann, D. (2013), Expressives and beyond, An introduction to varieties of use-conditional meaning, *Beyond expressives: Explorations in use-conditional meaning*, 1-58.

Gutzmann, D. (2015), *Use-conditional Meaning: Studies in Multidimensional Semantics*, Oxford, Oxford University Press.

H

Hare, R. M. (1952), *The language of morals* (No. 77), Oxford Paperbacks.

Hanks, P. (2011), Structured Propositions as types, *Mind*, Vol. 120, 477, jan. 2011.

Haslam, N. (2006), Dehumanization: An integrative review, *Personality and social psychology review*, 10(3), 252-264.

Haslanger, S. (2008), A social constructionist analysis of race, *Revisiting race in a genomic age*, 56-69.

Hay, R. J. (2013), Hybrid expressivism and the analogy between pejoratives and moral language, *European Journal of Philosophy*, 21(3), 450-474.

Hazlett, A. (2007), Grice's razor, *Metaphilosophy*, 38(5), 669-690.

Heim, I (1983), On the projection problem for presuppositions, In *Proceedings of WCCFL 2*, ed. D. Flickinger, 114–125, Stanford, CA: Stanford University Press.

Heim, I. (1992), Presupposition projection and the semantics of attitude verbs, *Journal of semantics*, *9*(3), 183-221.

Heim, I. (2004), Lectures notes on indexicality (Notes for class taught at MIT).

Hilbert, D. R. (1987), *Color and Color Perception*, Stanford, CA: CSLI Publications.

Holton, R. (1991), Intentions, response-dependence, and immunity from error, Response Dependent Concepts, Canberra: *ANU Working Papers in Philosophy*.

Hom, C. (2008), The Semantics of Racial Epithets, *Journal of Philosophy*, 105, 416-440.

Hom, C. (2010), Pejoratives, *Philosophy Compass*, 5(2), 164-185.

Hom, C. (2012), A puzzle about pejoratives, *Philosophical Studies*, 159, 383-405.

Hom, C. and May, R. (2013), Moral and Semantic Innocence, *Analytic Philosophy*, 54(3), 293-313.

Hom, C. and May, R. (2014), The inconsistency of the identity thesis, *Protosociology*, 31, 113-120.

Hom, C. and May, R. (forthcoming), Pejoratives as Fiction, in Sosa, D. (ed.) (forthcoming), *Bad words*, Oxford, Oxford University Press.

Horn, L. (1985), Metalinguistic Negation and Pragmatic Ambiguity, Language, 61(1), 121-174.

Horn, L. (1989), *A Natural History of Negation*, Chicago, University of Chicago Press.

Huang, Y. (2007), *Pragmatics*, Oxford: Oxford University Press.

Hume, D. (1739), *A treatise of human nature*, A. D. Lindsay (ed.) [1911], London: Dent.

Hurvich, L. M., Jameson, D. (1957), An Opponent-Process Theory of Color Vision, *Psychological Review*, 64 (6, Part I): 384-404.

IJ

Jackson, F. (1998), *From metaphysics to ethics: A defense of conceptual analysis*, Oxford University Press.

Jackson, F. and Pettit, P. (2002), Response–dependence without tears, *Noûs*, 36(s1):97–117.

Jahoda, G. (2015), *Images of savages: Ancient roots of modern prejudice in Western culture,* Routledge.

Jeshion, R. (2011), Dehumanizing slurs, In *Society for Exact Philosophy meeting*, Winnipeg, Manitoba.

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

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

Johnston,M.(1992), How to speak of the colors, *Philosophical Studies*,68:221.

Johnston, M. (2004), Subjectivism and "unmasking ", *Philosophy and Phenomenological Research*, LXIX(1):187–201.

Johnston, M., Smith, M., and Lewis, D. (1989), Dispositional theories of value, *Proceedings of the Aristotelian Society*, supplementary volumes, pages 89–174.

Joyce, R. (2009), Moral anti-realism, *The Stanford Encyclopedia of Philosophy*.

K

Kadmon, N. (2001), *Formal Pragmatics*, Oxford: Blackwell.

Kaplan, D. (1979), On the logic of demonstratives, *Journal of philosophical logic*, 8(1):81–98.

Kaplan, D. (1999), The meaning of ouch and oops: Explorations in the theory of meaning as use, Manuscript, UCLA. [Kaplan, D. (2001), The Meaning or Ouch and Oops (Explorations in the theory of Meaning as Use), Draft #3, ms., UCLA]

Karttunen, L. (1973), Presuppositions of compound sentences, *Linguistic inquiry*, pages 169–193.

Karttunen, L. (1974), Presupposition and linguistic context, *Theoretical Linguistics* 1:181–193.

Kelly, S. (2001a), Demonstrative concepts and experience, *Philosophical review*, 110:3:397-419.

Kirchin, S. (2013), *Thick concepts,* OUP Oxford.

Kirkland, S. L., Greenberg, J., & Pyszczynski, T. (1987), Further Evidence of the Deleterious Effects of Overheard Derogatory Ethnic Labels Derogation Beyond the Target, *Personality and Social Psychology Bulletin*, 13(2), 216-227.

Kitcher, P., (1999), Race, Ethnicity, Biology, Culture. *Racism*, 87-117.

L

Langton, R. (1998), *Kantian Humility: Our ignorance of things in themselves*, Oxford University Press.

Langton, R. (2012a), Beyond Belief: Pragmatics in Hate Speech and Pornography, in McGowan and Maitra (eds.), *What Speech Does*, Oxford, Oxford University Press.

Langton, R. (2012b), Language and Race, Speech and Harm: Controversies Over Free Speech.

Langton, R., Haslanger, S., Anderson, L. (2012), Language and Race, in Russell, G. and Graff Fara, D. (eds.), *Routledge Companion to the Philosophy of Language*, Routledge, 753-767.

Lamy, B. (1678), *L'art de parler, avec un discours dans lequel on donne une idée de l'art de persuader*, Troisième édition, Paris, chez André Pralard, ruë S. Jacques, à l'Occasion, Paris, 1676 (avec privilège du roy).

Lassiter, D. (2011), Vagueness as probabilistic linguistic knowledge, in Nouwen, R., Sauerland, U., Schmitz, H.-C., and van Rooij, R. (eds), *Vagueness in communication*, Springer Berlin Heidelberg, 127-150.

Leslie, S. J. (2013), Essence and Natural Kinds: When Science Meets Preschooler Intuition, *Oxford Studies in Epistemology*, 4(108), 9.

Leslie, S. J., (2015), "Hillary Clinton is the only man in the Obama administration": Dual Character Concepts, Generics, and Gender, *Analytic Philosophy*, 56(2), 111-141.

Levinson, S. C. (1983), *Pragmatics*, Cambridge: Cambridge University Press.

Levinson, S. C. (2000), *Presumptive Meanings: The Theory of Generalized Conversational Implicature*, Cambridge, MA: MIT Press.

Lewis, D. (1969), *Convention: A philosophical study,* Cambridge: Harvard University Press.

Lewis, D. (1979), Scorekeeping in a language game, *Journal of philosophical logic*, 8(1), 339-359.

Lewis, D. (1989), Dispositional theories of value, *Proceedings of the Aristotelian Society, Supplementary volumes*, 63, 113-137.

Locke, J. (1690), *An essay concerning human understanding*, London: Taylor [1722], 1.

M

Macià, J. (2002), Presuposicion y significado expressivo, *Theoria: Revista de Teoria, Historia y Fundamentos de la Ciencia*, 3 (45): 499-513.

Macià, J. (2006), Context, Presupposition and Expressive Meaning, Hand-out of a talk given at the Milan Meeting 2006.

Magri, B. (2009), A theory of individual-level predicates based on blind mandatory scalar implicatures, *Natural Language Semantics* 17:245–297.

Mates, B. (1969), *Synonymity*, Johnson Repring Corporation.

Matthen, M. (1988), Biological Function and Perceptual Content, *The Journal of Philosophy*, 95: 5-27.

Maund, B. (2012), Color, *The Stanford Encyclopedia of Philosophy*.

Mayr, E. (2002), The biology of race and the concept of equality, *Daedalus*, *131*(1), 89-94.

McCready, E. (2010), Varieties of conventional implicature, *Semantics and Pragmatics*, 3(8), 1-57.

McDowell, J. (1979), Virtue and Reason, *The Monist*, 62, 331-350.

McDowell, J. (1981), Non-cognitivism and Rule-Following, In Stephen Holtzman and Christopher Leich (eds.), *Wittgenstein: To Follow a Rule* (London: Routledge and Kegan Paul), 141-162.

McDowell, J. (1987), *Projection and Truth in Ethics*, Lindley Lecture, University of Kansas.

McDowell, J. (1994), *Mind and World*, Cambridge, MA: Harvard University Press.

McDowell, J. (1998), *Mind, Value, and Reality* (London: Harvard University Press).

Miller, A. (2012), Realism, *The Stanford Encyclopedia of Philosophy*.

Miller, W. I. (1998), *The anatomy of disgust*, Harvard University Press.

Millikan, R. G. (1984), *Language, Thought and Other Biological Categories*, Cambridge, MA: MIT Press.

Millikan, R. G. (1989a), In defense of proper functions, *Philosophy of science*, 56(2), 288-302.

Millikan, R. G. (1989b), Biosemantics, in *Journal of Philosophy*, 86: 281-97.

Miscevic, N. (2006), Moral Concepts: From Thickness to Response-Dependence, *Acta Analytica*, 21(1), 4-32.

Miscevic, N. (2011), Slurs & Thick Concepts - is the New Expressivism Tenable?, *Croatian Journal of Philosophy*, 32:159–182.

Miscevic, N. (2011b), No more tears in heaven: two views of response-dependence, *Acta Analytica*, 26(1), 75-93.

Morzycki, M. (2009), Degree modification of gradable nouns: size adjectives and adnominal degree morphemes, *Natural Language Semantics*, 17(2), 175-203.

Murdoch, I. (1956), Vision and choice in morality, *Proceedings of the Aristotelian Society, Supplementary Volume* 30: 32-58. Reprinted in her *Existentialists and Mystics*, Peter Conradi (ed.) (London: Chatto and Windus, 1997), 76-98.

Murdoch, I. (1957), Metaphysics and Ethics, in *The Nature of Metaphysics*, D. F. Pears (ed.) (London: Macmillan). Reprinted in her *Existentialists and Mystics*, Peter Conradi (ed.) (London: Chatto and Windus, 1997), 59-75.

Murdoch, I. (1962), The Idea of perfection, based on the Ballard Matthews Lecture delivered at the University College, North Wales. Reprinted in her *Existentialists and Mystics*, Peter Conradi (ed.) (London: Chatto and Windus, 1997), 299-336.

NO

Neufeld, E. (2018), An essentialist theory of the meaning of slurs, unpublished manuscript, USC.

Newman, G. E., Bloom, P., Knobe, J. (2014), Value judgments and the true self, *Personality and Social Psychology Bulletin*, 40(2), 203-216.

Nunberg, G. (2013), Slurs Aren't Special, draft.

Nunberg, G. (2017), The Social Life of Slurs, in Daniel Fogal, Daniel Harris, and Matt Moss (eds.) (2017): *New Work on Speech Acts*, Oxford: Oxford University Press.

O'Regan, J. K. and Noë, A. (2001), A sensorimotor account of vision and visual consciousness, *Behavioral and brain sciences*, 24(05):939–973.

P

Palmer, S. E. (1999), *Vision science: Photons to phenomenology*, The MIT press.

Panzeri, F. And Carrus, S. (2016), Slurs and Negation, *Phenomenology and Mind*, 11.

Peacocke, C. (1984), Colour concepts and colour experience, *Synthese*, 58(3), 365-381.

Peacocke, C. (1992), *A Study of Concepts*, The MIT Press.

Pelczar, M. and J. Rainsbury, (1998), The Indexical Character of Names, *Synthese* 114: 293-317.

Perry, J. (1993), *The problem of the essential indexical: and other essays*, Oxford University Press on Demand.

Pessoa, L. (2014), Précis of the cognitive-emotional brain, *Behavioral and brain sciences*, 38, p. e71.

Pettit, P. (1991), Realism and response-dependence, *Mind*, pages 587–626.

Pettit, P. (2003), Looks as powers, *Philosophical Issues*, 13(1):221–252.

Pitts, A. (2011), Exploring a 'Pragmatic Ambiguity' of Negation, *Language*, 87(2), 346-368.

Potts, C. (2003a), Expressive content as conventional implicatures, *Proceedings-Nels*, 33:303–322.

Potts, C. (2005), *The Logic of Conventional Implicatures*, Oxford University Press.

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

Predelli, S. (2013), *Meaning without truth*, Oxford University Press.

Putnam, H. (1975)[1970], *Is Semantics Possible? Mind, Language and Reality*, Cambridge: Cambridge University Press. 139–152.

Putnam, H. (1975), The meaning of "meaning", *Mind, Language and Reality*, Cambridge: Cambridge University Press. 215-271.

QR

Quine, W. V. (1960), *Word and Object*, Cambridge, Mass.

Randel Koons, J. (2003), Why response-dependence theories of morality are false, *Ethical Theory and Moral Practice*, 6(3), 275-294.

Recanati, F. (1981), *Les énoncés performatifs, contribution à la pragmatique*, Editions de Minuit.

Recanati, F. (2000), Deferential Concepts: A response to Woodfield, *Mind & Language*, 15(4), 452-464.

Recanati, F. (2001), Open quotation, *Mind*, 110(439), 637-687.

Recanati, F. (2002a), The Fodorian Fallacy, *Analysis*, 62(276), 285-289.

Recanati, F. (2002b), *Pragmatics and semantics*, Handbook of pragmatics.

Recanati, F. (2007), *Perspectival Thought: A Plea for (Moderate) Relativism*, Oxford University Press.

Recanati, F. (2013), Content, mood, and force, *Philosophy Compass*, 8(7):622– 632.

Reid, T. (1822) [1970], *An Inquiry into the Human Mind*, edited by T. Duggan, Chicago: Chicago University Press.

Rey-Debove, J. (1978), *Le métalangage, étude linguistique du discours sur le langage*, Paris, Le Robert, coll., L'ordre des mots.

Richard, M. (2008), *When Truth Gives Out*, Oxford University Press.

Risch, N., Burchard, E., Ziv, E., Tang, H. (2002), Categorization of humans in biomedical research: genes, race and disease, *Genome biology*, *3*(7).

Root, M. (2000), How we divide the world, *Philosophy of Science*, *67*, S628-S639.

Russell, B. (1905), On Denoting, *Mind*, 14, 479-493.

S

Sadock, J. M. (1978), On testing for conversational implicature, In P. Cole (Ed.), *Syntax and Semantics: Pragmatics*, 9 (pp. 281-298). New York: Academic Press.

Saka, P. (2000), Ought does not imply can, *American Philosophical Quarterly*, 37(2), 93-105.

Saka, P. (2007), Hate Speech, In *How to Think About Meaning* (pp. 121-153), Springer Netherlands.

Sainsbury, M. & Tye, M. (2013), *Seven Puzzles of Thought: And How to Solve Them: An Originalist Theory of Concepts*, Oxford University Press.

Sauerland, U. (2004), Scalar Implicatures in Complex Sentences, *Linguistics and Philosophy* 27(3), 367–391.

Sauerland, U. (2007), Beyond unpluggability, *Theoretical Linguistics*, 33(2), 231-236.

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

Schlenker, P. (2003), A plea for monsters, *Linguistics and philosophy*, 26(1), 29-120.

Schlenker, P. (2004), Context of thought and context of utterance: A note on free indirect discourse and the historical present, *Mind and Language*, 19(3), 279-304.

Schlenker, P. (2007), Expressive presuppositions, *Theoretical Linguistics*, 33(2), 237-245.

Schlenker, P. (2008), Be Articulate: A pragmatic theory of presupposition projection, *Theoretical Linguistics*, 34(3), 157-212.

Schlenker, P. (2011), The proviso problem: a note, *Natural language semantics*, *19*(4), 395-422.

Schlenker, P. (2015), Gesture projection and cosuppositions, Unpublished manuscript, Institut Jean Nicod and New York University.

Schlenker, p. (2016), The semantics/pragmatics interface, in M. Aloni and P. Dekker (eds), *The Cambridge Handbook of Formal Semantics*.

Schwarzschild, R. (1999), GIVENness, AvoidF and other constraints on the placement of accent, *Natural language semantics*, 7(2), 141-177.

Sayre-McCord, G. (2014), Metaethics, *The Stanford Encyclopedia of Philosophy*.

Simon, L., & Greenberg, J. (1996), Further progress in understanding the effects of derogatory ethnic labels: The role of preexisting attitudes toward the targeted group, *Personality and Social Psychology Bulletin*, 22(12), 1195-1204.

Simons, M. (2006), Foundational Issues in Presupposition, *Philosophy Compass*, 1, 357-72.

Simons, M., Tonhauser, J., Beaver, D., Roberts, C. (2010), What projects and why, In *Semantics and lintuistic theory* (Vol. 20, pp. 309-327)

Singh, R. (2008), *Modularity and locality in interpretation,* (Doctoral dissertation, Massachusetts Institute of Technology).

Soames, S. (1989), Presupposition, in *Handbook of Philosophical Logic*, Volume 4, ed. D. Gabbay and F. Guenthner (Dordrecht: Reidel), 553-616.

Soames, S. (2011), Propositions as cognitive event types, Chap. 6, In *New thinking about propositions*, by Jeff King, Scott Soames, Jeff Speaks, Oxford University Press.

Speaks, J. (2005), Is there a problem about nonconceptual content?, *The Philosophical Review*, 114(3), 359-398.

Spector, B. (2003), Scalar implicatures: Exhaustivity and Gricean reasoning, In B. ten Cate (Ed.), *Proceedings of the Eigth ESSLLI Student Session*, Vienna, Austria.

Stalnaker, R. (1970), Pragmatics, *Synthese*, 2(22):272–289.

Stalnaker, R. (1973), Presuppositions, *Journal of philosophical logic*, 2(4):447– 457.

Stalnaker, R. (1978), *Assertion*. Blackwell Publishers Ltd.

Strawson, P. F. (1950), On referring, *Mind* 59:320–344.

Strawson, P. F. (1964), Identifying reference and truth values, *Theoria* 30:96–118.

Stevenson, C. L. (1935), *The emotive meaning of ethical terms*, Harvard University Press.

Stevenson, C. L. (1944), *Ethics and Language*, New Haven: Yale University Press.

Stojanovic, I. (2016), Expressing aesthetic judgements in context, *Inquiry*, 59(6), 663-685.

Sundstrom, R. R. (2002), Race as a human kind, *Philosophy & social criticism*, *28*(1), 91-115.

T

Tarski, A. (1956), The concept of truth in formalized languages, *Logic, semantics, metamathematics*, 2, 152-278.

Taylor, P. C. (2013). *Race: A philosophical introduction*, Polity.

Thomason, R. (1990), Accommodation, Meaning, and Implicature: Interdisciplinary Foundations for Pragmatics, In *Intentions in Communication*, eds. Philip Cohen, Jerry Morgan and Martha Pollack, 325-363. Cambridge, MA: MIT Press.

Tirrell, L. (1999), "Racism, Sexism, and the Inferential Role Theory of Meaning", in *Language and Liberation: Feminism, Philosophy, and Language*, Hendricks, C. and Oliver, K., eds., Albany, NY: State University of New York Press, 41–79.

Tonhauser, J., Beaver, D., Roberts, C., Simons, M. (2013), Toward a taxonomy of projective content, *Language*, 89(1), 66-109.

UV

Vallentyne, P. (1996), Response-dependence, rigidification and objectivity, *Erkenntnis*, 44(1):101–112.

Van der Sandt, R. A. (1992), Presupposition projection as anaphora resolution, *Journal of semantics*, 9(4), 333-377.

VanBerkum, J. J., Holleman, B., Nieuwland, M., Otten, M., and Murre, J. (2009), Right or wrong? The brain's fast response to morally objectionable statements, *Psychological Science*, 20(9):1092–1099.

Von Fintel, K. (2004), Would you believe it? The King of France is back! Presuppositions and truth-value intuitions. In *Descriptions and beyond*, ed. Marga Reimer and Anne Bezuidenhout. Oxford University Press.

Von Fintel, K. (2012), Subjunctive conditionals, In Gillian Russell and Delia Graff Fara (Eds.), *The Routledge companion to philosophy of language*, 466-477, New York: Routledge.

Van Roojen, M. (2014), Moral cognitivism vs. Non-cognitivism, *The Stanford Encyclopedia of Philosophy*.

Vayrynen, P. (2013), *The Lewd, the Rude and the Nasty - A Study of Thick Concepts in Ethics*, Oxford, Oxford University Press.

Väyrynen, P. (2009), Objectionable thick concepts in denials, *Philosophical Perspectives*, 23(1), 439-469.

Väyrynen, P. (2012), Thick Concepts: Where's Evaluation?, *Oxford Studies in Metaethics*, 7, 235-70.

Väyrynen, P. (2013), Thick Concepts and Underdetermination, *Thick Concepts*, 136-60.

Von Frisch, K. (1950), *Bees*, Ithaca, NY: Cornell University Press.

W

Watkins, M. (2005), Seeing red, the metaphysics of colours without the physics, *Australasian Journal of Philosophy*, 83(1):33–52.

Wedgwood, R. (1998), The essence of response-dependence, *European Review of Philosophy*, 3:31–54.

Wiggins, D. (2016), *Continuants: their activity, their being, and their identity*, Oxford University Press.

Williams, B. (1985), *Ethics and the limits of Philosophy*, Harvard, Harvard University Press.

Williamson, T. (2003), Understanding and Inference, In *Aristotelian Society Supplementary Volume* (Vol. 77, No. 1, pp. 249-293), Oxford, UK: Oxford University Press.

Williamson, T. (2009), Reference, Inference, and the Semantics of Pejoratives, in *The Philosophy of David Kaplan*, Almog, J. and Leonardi, P., eds., Oxford: Oxford University Press, 137–158.

Wittgenstein, L., Anscombe, G. E. M., Wright, G. H., Paul, D., Bochner, M. (1969), *On certainty*, Blackwell Oxford, 174.

Wright, C. (1988), Moral values, projection, and secondary qualities, *Aristotelian Society Supplementary*, 73:1–26.

XYZ

Yablo, S. (2006), Non-catastrophic presupposition failure, *Content and Modality: Themes from the philosophy of Robert Stalnaker*, 164-190.

Zack, N. (2002), Philosophy of science and race, *Psychology Press*.

Zeevat, H. (1992), Presupposition and accommodation in update semantics, *Journal of semantics*, *9*(4), 379-412.

Résumé

Cette thèse s'intéresse à la structure, aux fonctions, et aux bases cognitives des termes d'offense (tels que le terme "boche"). Les termes d'offense, ainsi que leurs équivalents psychologiques, posent des problèmes intéressants et possiblement fondationnels à propos de la nature de la signification, de l'expressivité dans les langues naturelles, du rôle des émotions dans la catégorisation. Ce travail discute de ces questions - ainsi que de nombreuses autres - en s'intéressant à différentes théories existantes ou originales du phénomène. De nouvelles données linguistiques sont mises en avant qui remettent en cause des théories linguistiques telles que les visions vériconditionnelles ou présuppositionnelles du phénomène, et de nouvelles théories non-linguistiques du phénomène sont développées, invoquant les concepts de qualité seconde ou la notion d'essence. Les propriétés linguistiques particulières des termes d'offense, telles que la projection ou l'expressivité, apparaissent dans ce travail être des conséquences linguistiques d'un phénomène essentiellement psychologique : la possibilité d'une composante émotionnelle ou évaluative dans la structure même des concepts.

Abstract

The present work investigates the structure, function and cognitive underpinnings of slurring terms (such as "boche"). Slurring terms, and the mental correlates that I posit they have, raise interesting and possibly foundational issues about the nature of meaning, about expressivity in natural language, about the role of emotions in categorization. I discuss these questions - among many others - by studying different existing or original accounts of the phenomenon. I present novel linguistic evidence against linguistic views such as truth-conditional or presuppositional accounts, and develop new psychological (i.e. non-linguistic) theories of the phenomenon based on a connection with response-dependent concepts, or with essentialist concepts. The interesting linguistic properties of slurs, such as projection and expressivity, appear to be the linguistic consequences of the essentially mental fact that concepts may be loaded with emotional or evaluative content.

Mots Clés

Philosophie, Linguistique, Expressifs, Response-Dependence, Projection, Essentialisme

Keywords

Philosophy, Linguistics, Slurs, Response-Dependence, Projection, Essentialism