CROATIAN JOURNAL OF PHILOSOPHY

CROATIAN

JOURNAL

OF PHILOSOPHY

Vol. XVII · No. 50 · 2017

Book Reviews

Introduction

This is the 50th issue of the Croatian Journal of Philosophy. *The Journal, published since its first day by Kruno Zakarija, the best analytic publisher in Croatia, has been combining the best of the local work on analytic philosophy, done in Croatia and neighboring countries, Slovenia, Serbia and Hungary, and the cutting edge international work in the same area. The main cooperation has been, officially with the Institute of Philosophy in Zagreb, and not formally but very intensely with IUC Dubrovnik and the Department of philosophy in Rijeka. The high quality international conferences from IUC, dedicated to philosophy of language, philosophy of science and mathematics, philosophy of mind, ethics and metaphysics have been, for decade and a half, providing the precious framework of dialogue and cooperation, well documented in the thematic issues published throughout seventeen years.*

The present issue primarily belongs to the philosophy of language and linguistics series, supported by the homonymous IUC course. The preoccupation with philosophy of language has been a lasting feature of the local philosophy group, initially strongly inspired by the efforts of Georges Rey and Michael Devitt, beginning more than three decades ago, during their longer stay in Croatia. Other colleagues have then joined in, all the way to the present guest course directors, Barry Smith, Frances Egan, Michael Glanzberg and Jeff King. The present issue reflects the interests of a large part of the last year IUC meeting (the second part is coming out soon), with a focus on philosophy of pejoratives.

In her paper, "Loaded Words and Expressive Words: Assessing Two Semantic Frameworks for Slurs", Robin Jeshion *assesses the relative merits of two semantic frameworks for slurring terms. Each aims to distinguish slurs from their neutral counterparts via their semantics. On one, recently developed by Kent Bach, that which differentiates the slurring term from its neutral counterpart is encoded as a 'loaded' descriptive content. Whereas the neutral counterpart 'NC' references a group, the slur has as its content "NC, and therefore contemptible". On the other, a version of hybrid expressivism, the semantically encoded aspect of a slurring term that distinguishes it from its neutral counterpart is, rather, expressed. On this view, while the speaker's attitude may be evaluated for appropriateness, the expressivist component of slurring terms is truth-conditionally irrelevant. Jeshion argues that hybrid expressivism offers a more parsimonious analysis of slurs' projective behavior than loaded descriptivism and that its truth conditional semantics is not inferior to the possible accounts available for loaded descriptivism. She also meets*

Bach's important objection that hybrid expressivism cannot account for uses of slurring terms in indirect quotation and attitude attributions.

The book A Word Which Bears a Sword *(published in Zagreb by* Kru-Zak*) came out in 2016 and was discussed at the Philosophy of Linguistics and Language conference the same year in Dubrovnik. In the* "*Precis of the theoretical part of the book* A Word Which Bears a Sword" *Nenad Miščević presents his own view of pejoratives as negative terms for alleged social kinds: ethnic, gender, racial, and other. He argues that they manage to refer the way kind-terms do, relatively independently of false elements contained in their senses. This proposal, as presented in the book, is called Negative Hybrid Social Kind Term theory, or NHSKT theory, for short. The theory treats the content of pejoratives as unitary, in analogy with unitary thick concepts: both neutral-cum-negative properties (vices) ascribed and negative prescriptions voiced are part of the semantics preferably with some truth-conditional impact, and even the expression of attitudes is part of the semantic potential, although not necessarily the truth conditional one. Pejoratives are thus directly analogue to laudatives, and in matters of reference close to non-evaluative, e.g. superstitious social kind terms (names of zodiacal signs, or terms like* "*magician*")*. A pejorative sentence typically expresses more than one proposition and pragmatic context selects the relevant one. Some propositions expressed can be non-offensive and true, other, more typical, are offensive and false. Pejoratives are typically face attacking devices, although they might have other relevant uses. The NHSKT proposal thus fits quite well with leading theories of (im-)politeness, which can offer a fine account of their typical pragmatics.*

Testimonial injustice is a hot topic in social epistemology. In her contribution Julija Perhat *whose work is focused pejoratives (in particular, gender pejoratives for women) tries to connect them with injustice. Here she gives a precis of pejoratives and testimonial injustice and her present topic is testimonial injustice perpetrated by the serious use of pejoratives, in particular, gender pejoratives. Perhart combines two strands: on the one hand, the work on testimonial injustice where she relies on Miranda Fricker's work, and on the other hand, her own central area of interest, gender pejoratives.*

Katherine Ritchie *in her article* "*Social Identity, Indexicality, and the Appropriation of Slurs*" *stresses the point that slurs are expressions that can be used to demean and dehumanize targets based on their membership in racial, ethnic, religious, gender, or sexual orientation groups. Almost all treatments of slurs posit that they have derogatory content of some sort. Such views—which she calls content-based—must explain why in cases of appropriation slurs fail to express their standard derogatory contents. A popular strategy is to take appropriated slurs to be ambiguous; they have both a derogatory content and a positive appropriated content. However, if appropriated slurs are ambiguous, why can only members in the target group use them to express a non-offensive/posi-*

tive meaning? Here, she develops and motivates an answer that could be adopted by any content-based theorist. She argues that appropriated contents of slurs include a plural first-person pronoun. She shows how the semantics of pronouns like 'we' can be put to use to explain why only some can use a slur to express its appropriated content. Moreover, she argues that the picture she develops is motivated by the process of appropriation and helps to explain how it achieves its aims of promoting group solidarity and positive group identity.

Bianca Cepollaro *in her "Let's not worry about the reclamation worry" discusses the Reclamation Worry (RW), raised by Anderson and Lepore 2013 and addressed by Ritchie (this issue) concerning the appropriation of slurs. She argues that Ritchie's way to solve the RW is not adequate and she tries to show why such an apparent worry is not actually problematic and should not lead us to postulate a rich complex semantics for reclaimed slurs. To this end, after illustrating the phenomenon of appropriation of slurs, she introduces the Reclamation Worry, and then argues that Richie's complex proposal is not needed to explain the phenomenon. To show that, she compares the case of reclaimed and non-reclaimed slurs to the case of polysemic personal pronouns featuring, among others, in many Romance languages. She introduces the notion of 'authoritativeness' that she takes to be crucial to account for reclamation and focuses on particular cases (the "outsider" cases) that support her claims and speak against the parsimony of the indexical account. She concludes with a methodological remark about the ways in which the debate on appropriation has developed in the literature.*

Next three papers have a different theme. In her contribution "The Myth of Embodied Metaphor (the paper was also presented at the Philosophy of Linguistics and Language Conference in Dubrovnik 2016) Nikola Kompa *is critical of the leading embodied metaphor approach. She points out that according to a traditionally influential idea metaphors have mostly ornamental value. However, current research stresses the cognitive purposes metaphors serve. According to the Conceptual Theory of Metaphor (CTM) expressions are commonly used metaphorically in order to conceptualize abstract and mental phenomena. More specifically, proponents of CTM claim that abstract terms are understood by means of metaphors and that metaphor comprehension, in turn, is embodied. In this paper, Nikola Kompa argues that CTM fails on both counts.*

In contemporary epistemology, the view is that in order to have knowledge it is necessary to have an appropriately based belief. Guido Melchior *in his paper under the title of "Baseless Knowledge" argues that baseless knowledge can then be defined as knowledge where the belief is acquired and sustained in a way that does not track the truth. He argues that rejecting this view leads to controversial consequences but he does not say which belief bases constitute a sufficient condition for knowledge. The point he is making is that assuming that appropriate bases constitute a necessary condition for knowledge has controversial consequences.*

And finally, Adelin Costin Dumitru *in "On the Moral Irrelevance of a Global Basic Structure: Prospects for a Satisficing Sufficientarian Theory of Global Justice" interrogates the Rawlsian concept of a basic structure in the context of global justice. His aim in this paper is twofold: to show that the existence of a global basic structure is irrelevant from the standpoint of justice; and to set the stage for a cosmopolitan theory of global justice that employs satisficing sufficientarianism as a distributive principle.*

DUNJA JUTRONIĆ
NENAD MIŠČEVIĆ

Loaded Words and Expressive Words: Assessing Two Semantic Frameworks for Slurs

ROBIN JESHION*
University of Southern California, Los Angeles, USA

In this paper, I assess the relative merits of two semantic frameworks for slurring terms. Each aims to distinguish slurs from their neutral counterparts via their semantics. On one, recently developed by Kent Bach, that which differentiates the slurring term from its neutral counterpart is encoded as a 'loaded' descriptive content. Whereas the neutral counterpart 'NC' references a group, the slur has as its content "NC, and therefore contemptible". On the other, a version of hybrid expressivism, the semantically encoded aspect of a slurring term that distinguishes it from its neutral counterpart is, rather, expressed. A speaker who uses the slurring term references the group referenced by the neutral counterpart and, in addition, expresses her contempt for the target. On this view, while the speaker's attitude may be evaluated for appropriateness, the expressivist component of slurring terms is truth-conditionally irrelevant. The reference to the group, and only the reference to the group, contributes to truth conditions. I'll argue that hybrid expressivism offers a more parsimonious analysis of slurs' projective behavior than loaded descriptivism and that its truth conditional semantics is not inferior to the possible accounts available for loaded descriptivism. I also meet Bach's important objection that hybrid expressivism cannot account for uses of slurring terms in indirect quotation and attitude attributions.

Keywords: Slurs, pejoratives, semantics, expressivism, epithet.

In this paper, I assess the relative merits of two semantic frameworks for slurring terms. Each aims to distinguish slurs from their neutral counterparts via their semantics. On one, recently developed by Kent Bach (2017), that which differentiates the slurring term from its neutral counterpart is encoded as a 'loaded' descriptive content. Whereas the neutral counterpart 'NC' references a group, the slur has as its content "*NC, and therefore contemptible*". On the other, a version of hybrid expressivism, the semantically encoded aspect of a slurring term that distinguishes it from its neutral counterpart is, rather, *expressed*. A speaker who uses the slurring term references the group referenced by the neutral counterpart and, in addition, expresses her contempt for the target (the group and also possibly a particular individual in the group) on account of the target being an NC. On this view, while the speaker's attitude may be evaluated for appropriateness, the expressivist component of slurring terms is truth-conditionally irrelevant. The reference to the group, and only the reference to the group, contributes to truth conditions.

Bach offers his account as an improvement on reigning versions of semantic descriptivism, and as a competitor to hybrid expressivism, especially with respect to its ability to explain slurs' projective behavior. Hybrid expressivism offers a parsimonious analysis of slurs' projective behavior. Bach's account, I'll argue, is considerably less so.

One striking feature of Bach's account is that certain sentences containing slurs are neither true nor false and are so at least in part because of general purely linguistic features of sentences containing slurs. If cogent, the account has the resources to deliver the morally satisfying result that a sentence like "Jews are Kikes" is not true. Though Bach champions the result as a key advantage of his theory over hybrid expressivism, he does not spell out an overarching truth conditional semantics. I attempt to assess his view by constructing possible truth conditional semantic theories, and conclude that none delivers satisfying results without being *ad hoc*.

Finally, I attempt to meet three objections presented by Bach. One is that hybrid expressivism mistakenly gives priority to expressions of contempt over assertions of contemptability. Another is that it is unable to account for certain uses of slurring terms in indirect quotation and attitude attributions. The last is that it implausibly entails that all slurs and their neutral counterparts are co-extensive.

1. *Loaded Descriptivism: Projective Behavior*

Bach distinguishes two basic types of slurs, *group slurs* and *personal slurs*. For him, group slurs break down (roughly) into racial slurs ('chink', 'goy', 'kike', 'nigger', 'honkey'), political slurs ('commie', 'Nazi'), and religious slurs ('raghead', 'kike', 'heathen') while personal slurs encompass a wide range including those based on intelligence ('retard', 'moron'), character ('brown-noser', 'asshole'), sexuality ('queer', 'faggot',

'dyke', 'slut', 'lecher'), substance abuse ('boozer', 'acid freak'), and profession ('scab', 'pimp', 'whore').[1] His semantics applies across the board, to all these expressions.

Bach's novel theory differentiates slurs from their neutral counterparts in many ways that mirror that of Hom and May (2013). Like them, it differentiates semantically insofar as slurs and their neutral counterparts have distinct semantics, by which I mean that they are governed by different semantic conventions, different rules of use. It differentiates descriptively insofar as the distinguishing semantic feature of slurs is encoded as descriptive content. Yet Bach's motivations and handling of truth conditions mark important departures.

For Hom and May, 'kike' semantically encodes the descriptive content *ought to be the target of negative moral evaluation because of being Jewish*. Because it is a moral truth that no one ought to be the target of negative moral evaluation for being Jewish, on their view 'kike' has an empty extension. But not just 'kike'. "Pimp" and even "fucking Nazi"[2] have empty extensions. For them, slurs *as a class* have null extensions: having an empty extension is essential to what makes a slur a slur. In fact, one of their underlying motivations or initial uncontested data points is that all slurs necessarily have null extensions.

This, Bach rejects, and not just the point about necessity. On his view, whether a slurring term has an empty extension depends upon moral facts. It depends upon whether anyone is worthy of contempt on account of being an NC. For Bach, 'kike' includes both 'the property of being Jewish and the property of being contemptible in virtue of being Jewish' (Bach 2017: 6). So it is extensionless, while 'pimp' is not.

Bach advocates a distinctive way in which these two properties are encoded in slurs. In general, a slur has as its content "*NC, hence contemptible*". Thus, the meaning of a slur involves a categorizing component, equivalent with that of its neutral counterpart, conjoined with an additional evaluative component. The novelty in Bach's theory resides in how the semantic 'conjoining' operates. It appears to be explicitly designed to sidestep the problems created by Hom and May's handling of slurs' projective behavior.[3]

For Bach, 'kike' in

(1) Jake is a Kike

---

[1] I selected these examples from Bach's full appendix of slurs, including his typology. As Bach notes, many of the expressions have multiple uses and belong in more than one category (witness 'kike'). Even with this qualification, I do not endorse Bach's classification. Many of the expressions he categorizes as slurs, I regard as falling into different categories of pejoratives. I am here just adopting his typology to illustrate aspects of Loaded Descriptivism.

[2] Hom and May's PEJ operator ensures that 'fucking Nazi' counts as a slur. It is somewhat unclear if it does for Bach.

[3] I will not go into those problems here. I give an abbreviated account of the problems in the appendix to Jeshion (2013a), "Embracing Corruption: A Response to Hom and May". Sennet and Copp (2015) offer a comprehensive survey.

functions in *some* ways akin to the way that 'bachelor' functions, expressing the conjunctive property of being an unmarried male. Just as being a bachelor requires being both a male and being unmarried, so too, for Bach, being a Kike requires being both Jewish and being contemptible in virtue of being Jewish. (Bach 2017: 6) The extension of 'bachelor' contains all and only those individuals who are both male and unmarried, and thus is *true of* just those individuals.[4] Symmetrically, the extension of 'kike' is determined by whether there are individuals who are Jewish and contemptible by virtue of being Jewish. Because there are no such individuals, 'kike' has an empty extension. There is no one it is *true of*.

There is, however, a key difference with 'bachelor'. In

(2) Jake is a bachelor

the two properties of being male and being unmarried that determine the term's extension and the range of what it is true of operate together to contribute a single content. Thus, (2) encodes *a single proposition*, one that is essentially semantically equivalent with that expressed by

(3) Jake is an unmarried male.

By contrast, according to Bach, (1) encodes two distinct propositions, those expressed by

(4) Jake is Jewish

and

(5) Jake is contemptible in virtue of being Jewish,

which are truth conditionally independent of one another. So while 'bachelor' models the two conditions on being a kike, what determines the extension of 'kike', and what it is true of, it fails to model the way that slurs contribute to propositional contents of sentence.

To expose how slurs do so, Bach appeals to sentences involving non-restrictive relative clauses like

(6) Buffalo Bill, who was born in Buffalo, was a great showman.

(6) express two propositions, a primary *at-issue* content, expressed by

(7) Buffalo Bill was a great showman

and a secondary, supplementary content, expressed in the relative clause

(8) Buffalo Bill was born in Buffalo.

Importantly, (6) does not encode the single conjunctive proposition

(9) Buffalo Bill was a great showman and was born in Buffalo

that is true if and only if Buffalo Bill was a great showman and born in Buffalo. For Bach, this is the situation with slurs, only the two contents

---

[4] With 'bachelor', I am of course oversimplifying: unavailable unmarried men, say priests, don't count as bachelors. Nothing hangs here on the precise descriptive content of 'bachelor'.

are 'not given separate linguistic expression' (Bach, 7). Instead, what's expressed by (4) is primary, while that of (5) is secondary, functioning, he says, as a 'side comment' that is 'loaded into the slur'. What differentiates a slurring term from its neutral counterpart is the encoded 'loaded' descriptive content, hence the apt name *Loaded Descriptivism*.

Bach's semantic analysis of slurs is truly novel insofar as it posits that the meaning of a single term contributes twice, and separately, to the truth conditional propositional content of sentences in which it occurs. What justifies positing such semantic structure? Bach advances two reasons for modeling the contribution of slurs on analogy with sentences involving non-restrictive relative clauses. One concerns accounting for the projective behavior of the offensive element in slurs. As is well-known, what is 'offensive' in slurs projects out from many linguistic environments, occurring within the scope of negations and modals, and in the antecedents of conditionals in declaratives, as in (10) and (11), modeled on examples from Bach. In most instances, this 'offensive' element is *speaker-oriented* (Potts 2005).

(10) Jake is not a Kike. He only looks like one.

(11) If Jake is a Kike, he's stingy.

According to Bach, non-restrictive relative clauses exhibit the same projective behavior. He offers the following examples,

(12) It is not true that Buffalo Bill, who was born in Buffalo, was a great showman.

(13) If Buffalo Bill, who was born in Buffalo, was a great showman, he was popular.

noting that in (12) and (13), we rightly assume that the speaker asserts the secondary content encoded in the relative clause. Presumably, Bach would flesh out the dual contents of (10) and (11) thus, where I am representing the secondary speaker-oriented side comment parenthetically.

(14) Jake is not a Jew (Jews being contemptible in virtue of being Jewish). He only looks like one.

(15) If Jake is a Jew (hence contemptible), he is stingy.

Sentences involving slurs in subject rather than predicate position would also receive the dual contents. So on loaded descriptivism

(16) Kikes don't celebrate Easter.

is semantically equivalent to

(17) Jews, who are contemptible on account of being Jewish, don't celebrate Easter.

According to loaded descriptivism, all declarative sentences with a slur in either subject or predicate position will encode both primary and secondary contents. I will say that such sentences possess a *dual proposition structure*.

An initial difficulty is that the projective behavior of slurs ranges more widely than these dual proposition structure examples reveal. What is offensive in slurs projects not only in complex declarative constructions but as well in interrogatives, imperatives, and vocatives:

(18) Why were you talking to those Kikes?
(19) Stay away from those Kikes.
(20) Kike!

To maintain a uniform analysis to handle the projective behavior, Bach's account requires introducing a secondary contribution of slurs in addition to that which they contribute to the encoded non-declarative. Keeping parity with (10) and (11), the slurs in (18)–(20) ought to contribute a speaker-asserted side comment of a declarative content along with the primary content in the interrogative, imperative, vocative:

(21) Why are you talking to those Jews (who are contemptible in virtue of being Jewish)?
(22) Stay away from those Jews (who are contemptible in virtue of being Jewish).
(23) Jew (hence contemptible)![5]

This awkward result highlights the *prima facie* implausibility of appealing to a distinct secondary content to explain the projective behavior of slurs. After all, the speakers in (18)–(20) do not seem to be encoding declarative contents *at all*.

Another problem is that a sentence like "Kikes are contemptible" ought to strike us as tautological or analytic truths. But this is far from clear. They do not elicit the "duh, you called them *Kikes*!" response in the way that "Kikes are Jews" does.

Bach's second reason for appealing to the semantics of non-restrictive relative clauses concerns parallels in their discourse denial structure. Suppose your interlocutor assertively utters (6) and you know that while Buffalo Bill was indeed a great showman, he was not born in Buffalo. Wishing to register a correction is knotty because a blanket assertion of "that's not true" would fail doubly: it would naturally be understood as a rejection of the primary content, which you endorse, and would not be understood as a rejection of the secondary content, which you deny. Though the sentence encodes two distinct propositions, given its syntactic structure, your 'that' is naturally construed as referencing just its primary content. Now suppose your interlocutor assertively utters (1), and you know that Jake is Jewish. Wishing to issue a denial of the smearing element, "that's not true" will again fail doubly: it will naturally be taken as a denial that Jake is not Jewish, not anything

[5] Notice that the secondary content cannot be within the scope of the question. Such a move would misrepresent the bigot's utterance. One who asks (18) is not inquiring why you are talking to those Jews that are contemptible in virtue of being Jewish. Someone who asks "Is there a Chink in the kitchen?" is not making the bizarre yet seemingly innocent inquiry whether there is in the kitchen anyone who is Chinese and contemptible on account of being Chinese.

else possibly encoded in the slur. To Bach, the striking parallels offers strong justification for modeling the semantics of slurs on sentences containing non-restrictive relative clauses.

This rationale likewise raises questions. True, bare, blanket denials with a simple "no" or "that's not true" will not isolate the smearing element in the slur-utterance, just as they do not with the content of the relative clause. The reason, though, is because there exists an independent descriptive content in both sentences naturally regarded as the default object of such denials. It does not *support* accounting for the smearing element of the slur with a secondary descriptive content. In fact, pressing further on the comparison exposes important disanalogies.

For one, the content in non-restrictive relative clauses are amenable to denial with a more specific negation-containing denial, say, "No, he was not born in Buffalo" or "Well, no, he was not born in Buffalo", with 'well' functioning as an acknowledgement of the truth in the primary content (7). This is possible because the existence and distinctness of the two contents is represented syntactically and semantically. If Bach's comparison is apt, then, a denial like "Well, no, Jews are not contemptible on account of being Jewish" to (1), as well as to (10), (11), (14), (15), should immediately strike us as the right sort to issue to the smearing element in the slur. But this is far from apparent. Certainly the form of the denial suggests that the speaker has simply advanced an incorrect belief about Jews, not that they have *done something* derogating in using the slur. True, Bach could maintain that this isn't the only sort of push-back a recalcitrant hearer would want to issue. The speaker would additionally be open to censure for the performative, to what he *did* in using the slur. But if Loaded Descriptivism is correct, the "Jews are not contemptable" denial should nevertheless still strike us as obviously apt, and it isn't.[6]

Another disanalogy concerns the fact that, for sentences containing slurs, there is no special reason why reference to the group, the alleged primary content according to Loaded Descriptivism, is in fact *primary* in the sense of being that which is the subject of discourse, the possible subject of dispute. With sentences containing non-restrictive relative clauses, the syntactic form itself reveals its primary content *as* primary. Not so for slur-containing sentences, where there is no syntactic representation of the two distinct contents at all. The point and its significance emerge when we consider contexts in which the group membership is not at all at issue in the sense of being a possible subject

---

[6] In addressing objections regarding discourse denials, Bach makes a curious claim at odds with the loaded descriptivist semantics. He claims that someone who uses 'kike' in a sentence like (1) does not *assert* but only *presupposes* that Jews are contemptible on account of being Jewish. (Bach 2017: 14). If the semantics are given as Bach initially details them, with 'kike' contributing an additional descriptive 'side comment' modeled on the semantics of sentences with non-restrictive relative clauses, the speaker must be making an assertion. Presuppositional accounts of slurs have a wholly different semantic structure.

of dispute to the discussants, while the smearing element is. Imagine neo-Nazis showing up at a public Jewish parade, saying

(24) The Kikes are marching down Fifth Avenue.

Here, everyone knows—and knows that everyone knows—that those he is pointing to are Jews and that they are marching. Only the smearing element of the slur is salient. Thus, an utterance of "no" or "that's not true" ought to seem obviously *apropos* as a direct denial of the alleged asserted content that Jews are contemptible in virtue of being Jewish – for in this context, it has no competitor-contents. But it does not.

Similarly, "no" or "that's not true" should seem apt in reply to a question like (21) and an imperative like (22). If they contain a side-comment declarative content, as I argue above loaded descriptivism ought to countenance to preserve uniformity in handling the smearing element, such denials should seem on target. In fact, they should seem especially apt in the absence of competitor declarative contents. But they do not. They come off as strange.

2. *Loaded Descriptivism: Truth Conditional Semantics*

Where does this leave us with respect to the truth value of sentences containing slurs? Certainly positing dual proposition structure to slur-containing declarative sentences will complicate their overall truth-conditional semantics. Bach regards the ensuing complications as one of the theory's strengths. Indeed, he champions his loaded descriptivist semantics as providing a novel and attractive illustration why we 'shouldn't have to decide' on the truth of sentences like (1). Speaking of that sentence, Bach writes:

> Is this true or false? Is it or is it not the case that Jacob is a Kike? On the one hand, you might say, "yes, he is a Kike", since the word 'kike', notwithstanding its derogatory force, does manage to distinguish Jews from non-Jews. On the other hand, you might say, "No, though Jewish, he is not a Kike" (perhaps because you agree with me that being a Kike requires being contemptible for being Jewish). In the recent debate about slurs, some lean one way, some the other. In my view, one shouldn't have to decide – having to say 'yes' or 'no' to a statement made with a sentence like (1) is a forced choice that one should resist making.
>
> …according to Loaded Descriptivism, the problem with assessing (1) for truth or falsity, and why we resist doing this, is its misleadingly simple predicative form. Just recall our non-semantic question...: what's the difference, if any, between being a Jew and being a Kike? From the perspective of Loaded Descriptivism, this is a misguided question. On the one hand, one just has to be a Jew; on the other, being a Kike requires that being Jewish inherently involves being contemptible. Since there are actually two separate propositions to be considered, it is a mistake to suppose that a sentence like (1) has a single truth-value. Like a sentence containing a nonrestrictive relative clause but in a compressed way, it expresses two independent propositions, not one conjunctive proposition. (Bach 2017: 7–8)

Bach's thought appears to be that the reason why a sentence like (1) lacks truth value is that it possesses a dual proposition structure. Bach concludes that this 'undercuts' debates extending from Hom and May's (2013) on the truth values of sentences containing slurs, and even on whether slurs like 'kike' have a null or non-null extension (Bach 2017: 7–8).

Things are trickier than they might appear, however. Bach frames his 'forced-choice' point specifically about sentences like (1), yet advances loaded descriptivism's semantics to apply *generally*, to all slur-containing sentences. What remains unclear, then, is how this impacts loaded descriptivism's truth-conditional semantics. To evaluate the cogency of the theory and plausibility of construing (1) as lacking truth-value, we need answers to the following: What is the underlying rationale for the 'forced-choice' claim about (1)? Do all slur-containing sentences fail to have a single truth value, or only those sharing additional structure to (1)? What are the truth conditions for all slur-containing sentences? [7]

Bach does not directly address these questions. He says enough, however, to steer us toward constructing the most plausible answers. My goals here are partly clarificatory, partly interpretive, and partly critical. I'll first detail the space of options for loaded descriptivism's truth-conditional semantics. I'll then assess them both interpretively, according to their capacity to account for Bach's professed commitments, and philosophically.

As I see it, based on his remarks above, Bach has the following options for the truth conditional semantics of loaded descriptivism:

Conjunction Theory: A slur-containing sentence with dual proposition structure is true if and only if its primary and secondary contents are true. Otherwise it is false.

Total Truth-Value Gap Theory: All slur-containing sentences with dual proposition structure are neither true nor false.

Symmetric Truth-Value Gap Theory: Slurs-containing sentences with dual proposition structure lack truth value if and only the primary and secondary contents come apart. Where they are the same, the truth value of the sentence is conjunctive.

Primary Dominant Truth-Value Gap Theory: Slur-containing sentences with dual proposition structure lack truth value if and only if the primary content is true and the secondary content is false. Otherwise the truth value of the sentence is conjunctive.

Let's illustrate the truth-conditional semantics of each option for four sentences having the same syntactic form as (1), with the slurs all oc-

---

[7] Ideally, we also would want an account of how loaded descriptivism's truth-conditional semantics impacts non-declaratives like interrogative and imperatives, where uniformity requires (I argued), in addition to the question and instruction, a secondary assertive content. Do they too have a truth value for their assertive content?

cupying the same predicate position. Assume that no one is worthy of contempt on account of religious affiliation or ethnicity, but that exploitatively selling the sexual services of women and children makes one worthy of contempt. Assume also that Jake is Jewish, Blake is not, Jayden exploitatively sells the sexual services of women and children, while Brayden does not. Then the truth-values of the individual primary and secondary contents of each sentence is as given below.

(1)   Jake is a Kike.   Primary: T; Secondary: F
(25) Blake is a Kike.   Primary: F; Secondary: F
(26) Jayden is a pimp.   Primary: T; Secondary: T
(27) Brayden is a pimp.   Primary: F; Secondary: T

We can summarize how our candidate theories deliver the truth conditions for each sentence thus:

Conjunction Theory: (26) is true. (1), (25), (27) are false.

Total Truth-Value Gap Theory: (1), (25), (26), (27) are truth-valueless.

Symmetric Truth-Value Gap Theory: (26) is true, (25) is false. (1), (27) are truth-valueless.

Primary Dominant Truth-Value Gap Theory: (26) is true. (25), (27) are false. (1) is truth-valueless.

Now, which theory best squares with Bach's claims about loaded descriptivism? Clearly, Bach's invocation of the 'forced-choice' rationale on (1) marks a definite dismissal of the Conjunction Theory. He will reject any theory on which (1) possesses a determinate truth value.

Does this mean he's committed to the Total Truth-Value Gap Theory? There are good reasons to think so. His reference to 'misleadingly simple predictive form' in the quote above suggests that his rationale for saying (1) lacks truth value is simply its dual proposition structure alone. The theory has the virtue of being non-*ad-hoc*, an attractively general account of why (1) lacks truth-value. The interpretive problem is that Bach never explicitly signs onto a Total Truth-Value Gap theory. Moreover, he makes striking claims suggesting he rejects it. He states outright that loaded descriptivism

> does not entail that slurs are true of their targets or that they are not. In fact, it correctly allows that some slurs can be, and indeed are, true of their targets. For example, an asshole, in virtue of what makes him qualify as such, really is contemptible. Calling him an asshole may be rude or crude, but you don't misrepresent him by calling him that…What makes an asshole an asshole makes him contemptible. Nothing makes a Jew a kike, regardless of what anti-Semites may think. Whereas group slurs generally misrepresent their targets, many personal slurs often represent their targets accurately, however rudely. (Bach 2017: 13–14)

Bach is here underscoring that loaded descriptivism, unlike Hom and May's descriptivist semantics, leaves open whether slurs, *as slurs*, inherently misrepresent their targets or have an empty or nonempty extension.

Bach assures us that the world is populated with assholes. Now, if you say to an asshole

(28) You are an asshole

have you spoken truly? Is (28) true? This he does not explicitly pronounce on. Yet, we're told, 'asshole' has a non-empty extension and is sometimes true of its targets. And you have engaged in no misrepresentation in calling an asshole an asshole. It seems highly implausible, then, that for Bach (28) should count as truth-valueless.

To Bach, 'asshole' possesses exactly the same semantic structure as 'kike', a point I'd dispute. But let us bypass this and substitute a term that, we agree, is a slur, and has a non-empty extension – 'pimp'. If (28) is true on loaded descriptivism, then, given the assumptions sketched above, (26) is too. If we can speak truly in calling an asshole an asshole, we can speak truly in calling a pimp a pimp. When primary and secondary contents are both true, the whole is true and thus the Total Truth Value Gap Theory is ruled out as codifying the truth conditions of Bach's loaded descriptivism. This demonstrates that the dual proposition structure of slurs cannot be Bach's sole reason for denying slur-containing sentences such as (1) a truth value.

I've been arguing along an interpretative dimension: that Bach himself appears unlikely to accept total truth value gaps given that he allowed that certain slurs like 'asshole' and 'pimp' have non-empty extensions, are *true of* certain individuals, and that sentences containing them involve no misrepresentations of such individuals. Now I wish to ask, *should* Bach embrace the Total Truth Value-Gap Theory? Is it a viable truth conditional semantics for loaded descriptivism?

The chief problem stems from allowing that slurs have extensions and can be true of individuals yet every sentence containing a slur lacks truth value. A slur can be true of a person P yet we're unable to express that, truly, by *using*, as opposed to *mentioning*, it. While 'pimp', we suppose, is true of P, one cannot express that, truly, by saying "P is a pimp". A lawyer can speak truly, albeit pedantically, by saying "I have evidence to convict this person who 'pimp' is true of", but not by saying "I have evidence to convict the pimp". This, I submit, is not a happy consequence.

Let's turn then to the Symmetric Truth Value Gap Theory. It can be seen as resting on a slightly different rationale for denying (1) truth value: not dual proposition structure alone but such structure together with a conflict in primary and secondary content truth values. This too would be a non-*ad hoc* rationale, rooted exclusively in features of the (posited) semantic structure of slurs. It also has the virtue of accounting for Bach's appeal to a 'forced false choice' between a true and false sub-content. To evaluate the Symmetric Truth Value Gap Theory's candidacy, consider not just (1) but also sentences like (27) with primary content false, secondary true. Surely Bach should judge them false. If you can speak truly in calling an asshole an asshole you can speak falsely in calling a saint an asshole. And so long as (26) counts as true, (27) should certainly count as false, not as truth-valueless. 'Pimp' has

an extension and the speaker of (27) is just mistaken about Brayden's occupation. Is there any reason to say that we are confronting a misbegotten forced false choice? There is none. What was said was *false* solely in virtue of the fact that Brayden is not in that business. If this is correct, the Symmetric Truth-Value Gap Theory is ruled out, (1)'s truth value gap cannot be solely grounded on a conflict between primary and secondary contents' truth values.

This leaves us with the Primary Dominant Truth-Value Gap Theory which offers a conjunctive account of truth conditions for all sentences except those like (1) with primary content true, secondary false, which are truth-valueless. The philosophical problem with this theory is that it appears thoroughly *ad hoc*. The theory isolates sentences like (1) as special, with no underlying rationale for why they, and only they, lack truth value. It is also a poor interpretive analysis, failing to cohere with Bach's claim that the 'forced choice' in (1) is rooted in general features of its dual proposition structure.

Whether loaded descriptivism offers an improved descriptivist analysis of slurs turns in part on the extent to which it delivers a plausible and clearly motivated truth conditional semantics. We've considered four accounts. The Conjunction Theory, Total Truth Value Gap Theory, and Symmetrical Truth Value Gap Theory we've ruled out on interpretive grounds. The Conjunction Theory is plainly incompatible with Bach's claim that (1) lacks truth value. The Total Truth Value Gap Theory and the Symmetrical Truth Value Gap Theory appears at odds with Bach's commitment to certain slurs having extensions, being true of certain individuals. The Total Truth Value Gap Theory is also implausible on philosophical grounds for artificially barring slurs from encoding truths (or falsehoods) while allowing that some slurs are *true of* individuals. The Primary Dominant Truth Value Gap Theory is *ad hoc*. None looks wholly unproblematic.

Finally, let's have a look at how the non-conjunctive theories evaluate sentences like:

(29) Jews are Kikes   Primary: True, Secondary: False
(30) Jews are not Kikes.   Primary: False, Secondary: False
(31) Exploiters of women and
     children's sex are pimps. Primary: True, Secondary: True
(32) Exploiters of women and
     children's sex are not pimps. Primary: False, Secondary: True

Mirroring the patterning above, the Total Truth Value Gap Theory takes all to be truthvalueless. Consequently, (29) is classified alongside (30) and (31) in lacking truth value, which is hardly a morally satisfying result, at least not one that the theory could advertise as an explanatory virtue. The Symmetrical Truth Value Gap Theory groups (29) along with (32) as truthvalueless, which seems random. The Primary Dominant Truth Value Gap Theory offers a more satisfying result, with (29) truthvalueless, (31) true, (30) and (32) both false, yet, as noted earlier, stands

in need of a linguistic justification. On any of these theories, it is difficult to see how Loaded Descriptivism marks a clear improvement on the truth conditional semantics of competitor semantic descriptivist views.

3. *Hybrid Expressivism: Projective Behavior*

Hybrid expressivism differentiates between slurs and their neutral counterparts by incorporating an independent non-truth conditionally relevant semantic component, the expression of contempt toward the target group. On my favored view,[8] slurs function semantically in the same way that their neutral counterparts function when given contemptuous intonation and when fronted by expletives or certain negative adjectives. (1) receives roughly the same semantic analysis as utterances of (33) and (34).[9]

(33) Jake is a dirty Jew.
(34) Jake is a Jew^C.

'Kike' has a group-referencing component, picking out the same group as its neutral counterpart, 'Jew'. It is this component, and only this component, that contributes to determining the truth conditions of sentences it occurs within. (1), (33), and (34) are all truth conditionally equivalent to "Jake is a Jew".

Unsurprisingly, then, 'kike' has the same extension as 'Jew'. Some regard this as a fatal flaw of the theory. But if one recognizes that pejoration, especially the distinctive type of pejoration manifested in slurs, is also pulled off by intonationally marking the neutral counterpart with expressions of contempt or attaching contempt-flagging adjectives like 'dirty' to neutral counterparts, the impetus to immediately discredit hybrid expressivism drops away. For it is far from clear that by marking 'Jew' with contempt, one has shifted its referent.

'Kike' differs from 'Jew' insofar as it semantically encodes speaker-contempt. The semantics posits an expressive component in the form of a rule that 'kike' be used to express one's contempt toward Jews on account of being Jewish. Slurs are thereby classified (in part) alongside other expressives like 'yahoo', whose semantics is also given by a rule (very roughly) to use it to express pleasure in a significant event. Thus, the expression of contempt should not be assimilated to any kind of descriptive meaning, assertion, or presupposition that Jews are contemptable on account being Jewish.

---

[8] I have fleshed out the semantics and pragmatics of my own version of hybrid expressivism in Jeshion (2013b), (2016), (2017), (ms). Though the account differs in some respects from those developed by others, here I introduce only those features common to all. Other expressivists include Kaplan (2005), Saka (2007), Copp (2001), Potts (2005), (2007), Gutzman (2013). Richard (2008) offers a broadly expressivist analysis yet denies that slurs are truth conditionally equivalent to their neutral counterparts.

[9] There are some differences in the pragmatic analysis of how they cause offense. But I do not take these up here.

One immediate attractive consequence of hybrid expressivism is that it makes sense of the non-tautologousness of "Kikes are contemptible". This follows from the fact that contempt is encoded as an attitude, not descriptively. True, the speaker both expresses and predicates contempt to Jews. But that doesn't make its assertive content tautological. Compare: "Ouch, that hurts!" is not tautological though "Painful things hurt" is.[10]

Because it fully detaches the group-referencing component from the expressive component, hybrid expressivism offers an attractive account of the projective behavior of slurs. Both (10) and (11)

(10) Jake is not a Kike. He only looks like one.
(11) If Jake is a Kike, he's stingy.

encode exactly the same expression of contempt which projects out of the negation and modal. Loaded descriptivism, we saw, introduces its secondary content "Jews being contemptible on account of being Jewish" to handle the projection in these declaratives. It ran into problems, however, once we widened the scope of the projection behavior, so that uniform treatment implausibly required that an assertion be coupled together with questions, commands, calls. Hybrid expressivism offers a far more parsimonious treatment, smoothly accounting for (18)–(20)

(18) Why were you talking to those Kikes?
(19) Stay away from those Kikes.
(20) Kike!

by appeal to the expression of contempt, requiring no additional asserted content at all.

4. *Hybrid Expressivism: Indirect Quotation, Attitude Attributions, and Truth*

Bach advances three main problems for Hybrid Expressivism. The first concerns expressivism's account of the "import of slurs". He claims it gets the order of explanation backwards:

> Using a slur expresses contempt….not as a matter of meaning but because it imputes contemptibility to members of the target group. Yes, there is a big difference between calling someone a Jew and calling them a kike, but the difference consists in what is imputed (contemptability) and only derivatively in what (contempt) is thereby expressed. (Bach 2017: 10)

The thought is that expressions of contempt from the act of using a slurs are derivative, accounted for by reference to the encoded semantic content of being contemptible.

I find this intuitive judgment of explanatory priority perplexing. Proponents of loaded descriptivism and hybrid expressivism will large-

---

[10] Expressively encoded content may be reinforced without redundancy "That fucking fucker is such a fuck" expresses a heightening of the speaker's emotional state. Cf. Potts (2007).

ly agree that, typically, in using slurs speakers convey contempt *and* impute contemptability. But I see no way to establish, pre-theoretically, that contemptability is more primitive than the expression of contempt, or the reverse.

Embedded in this claim about which is more primitive may be a concern that hybrid expressivism cannot explain how speakers convey contemptability via the expression of their own contempt. The idea is that the expression of a mere subjective emotion could not elicit or impute the objective assessment of contemptability. Yet this misconstrues the nature of contempt. Contempt is an affective attitude, an emotive stance, that ranks its objects as lesser persons *qua* persons, relative to *interpersonally* shared moral norms. Because these norms are inescapably binding, by expressing one's own contempt with a slur, one effectively represents the person or group as contemptable.[11]

The second problem Bach isolates concerns slurs' behavior in attitude attributions. He claims that expressivist theories are unable to account for how incorporating slurs into attitude attributions can add to their accuracy. I illuminate the alleged trouble by first examining instances of indirect quotation and later confront attitude attributions. Suppose that Jen said

(1) Jake is a Kike

and I report on what she said with this instance of indirect quotation

(35) Jen said that Jake is a Kike.

With (35), the slur enables me to convey Jen's anti-Semitic attitudes about Jews and does not entail any such attitudes of my own. There's nothing infelicitous with my following up the report with "I was so disgusted". According to Bach, because expressivism "predicts that expressiveness always scopes out of embedded contexts", it "is not equipped to account for all that is being reported." (Bach 2017: 11) Expressivism appears to get things doubly wrong: because the semantics requires the encoding of the speaker's own contempt, (35) communicates an anti-Semitic attitude to me; and because it does not encode descriptive content of being contemptible, the report fails to attribute to Jen the anti-Semitic attitude she communicated with (1). Loaded Descriptivism looks far better on this score, for my report semantically encodes only something about Jen, that she said that Jake is a Jew, hence contemptible.

The argument proves too much. For if this is a good argument against a hybrid expressivist theory of slurs, it is a good argument against an expressivist semantics for any term at all, including those like 'bloody', 'freak'n', 'fucking', and 'goddamned' that are widely re-

---

[11] See Mason (2003) for a rich analysis of the structure of contempt, including how it is governed by interpersonal moral norms. Cf., also Bell (2013). I offer a fuller explanation of why expressions of contempt convey contemptability in Jeshion (2017).

garded as *requiring* expressivist treatment for their uses as intensifiers.[12] Suppose that Jen said

(36) The cats are terrified of that goddamned dog.

Then, by exactly the same argument, when I accurately report what Jen said with this instance of indirect quotation,

(37) Jen said that the cats are terrified of that goddamned dog

I am only able to express my own attitudes toward the dogs, nothing about Jen's. But this is implausible. Following up (37) with "But I really love that dog, I don't know why she hates it so" is not infelicitous.

Whatever complexities arise from specifying the semantics of bare expressives in indirect quotation shouldn't make us question an expressivist semantics when these terms occur in unembedded sentences.

I don't have the space here to offer a full account of the semantic contribution of slurring terms in indirect quotation and attitude attributions, but I will say enough to turn back this objection. When expressives occur in the embedded clause of indirect quotation, as in (35) and (37), the sentence alone typically admits both a *speaker-oriented* and an *attributee-oriented interpretation*. In the former, with the slur or the bare expressive, the speaker expresses her own attitudes, in the latter the speaker captures the attitudes of the one whose words she is reporting. In taking up Bach's challenge, we have been considering interpretations in which the attitudes expressed with 'kike' in (35) and 'goddamned' in (37) are exclusively attributee-oriented. But the sentences readily admit speaker-oriented interpretations as well. (35) could be felicitously followed up with "But of course that PC-queen Jen used 'Jew'", (37) with "I have no idea why she continues to love that fucking dog."

For speaker-oriented interpretations, the expressive component of slurs' meaning is governed by the standard rule to use it to express one's contempt. Yet within the indirect report, the group-referencing and expressive components diverge in which point of view they capture. The group-referencing component of 'kike' encodes what the attributee said, that Jake is a Jew, and the expressive component enables the speaker to express her own contempt toward Jews. The bare expressive in (37) functions similarly. For attributee-oriented interpretations, the group-referencing and expressive components of the slur align, capturing both what the attributee said and how she said it. Yet precisely because expressives standardly encode the attitudes of the speaker, indirect reports involving them require that the occurrence of the term be treated quotatively, as an instance of mixed quotation, where a *part* of the indirect report is construed quotatively. In (35), the whole slur or even just its expressive component can be understood as within quotes.

---

[12] See Potts (2005), (2007), Gutzman (2013).

This correctly functions to insulate the speaker from her report being taken as an expression of her own attitudes.[13]

Bach acknowledges the possibility of appealing to mixed quotation to explain the occurrences of slurs within indirect quotation. He locates the main problem as one involving attitude reports, illustrating the phenomena with a locution involving 'thought' as the verb in the attitude ascription and an 'according to…' operator:

(38) Dick thought that Henry was a Kraut and Zbig was a Pollack
(39) According to Dick, Henry was a Kraut and Zbig was a Pollack.

Here, in contrast with our examples of indirect quotation, the sentences themselves much more strongly suggest an attributee-oriented interpretation as the default. (39) in particular seems even to resist a speaker-oriented interpretation. The slurs encode how Dick thinks of Henry and Zbig, not how the speaker does. Hybrid expressivism, Bach tells us, is not equipped to account for what is being reported.

But it can. To see why, notice first that the primacy of the attributee-oriented interpretation is not a general feature of propositional attitude ascriptions. Indeed, it is noteworthy that Bach did *not* attempt to illustrate the problem with belief ascriptions like

(40) Dick believes that Henry is a Kraut and Zbig is a Pollack

which readily admits a speaker-oriented interpretation, attributing to Dick only beliefs that Henry is Jewish, Zbig Polish. Sentences containing other propositional attitude verbs – 'knows', for instance—also naturally admit speaker-oriented interpretations. The fact that propositional attitude ascriptions do not generally default to attributee-oriented interpretations is a tip-off that there is something special going on in (38) and (39).

What distinguishes (38) and especially (39) is that they bring us inside the internal mindset—the point of view—of Dick. They do so in a way reminiscent of the devices of free indirect discourse, and I will propose treating (38) and (39) in a way that is parasitic on how slurs in free indirect discourse operate.

Free indirect discourse is a third-person narrative form in which aspects of the perspective, voice, of a character are presented within the narration itself. Though it functions to capture the inner life of the character in the way that ordinary indirect speech does, it does not rely on phrases like 'he thought' and 'she wondered' to do so. Just as indirect quotation can be construed as a linguistic device to encode

---

[13] There is nothing *ad hoc* in this account. Mixed quotation is pervasive, a phenomenon we need to explain other varieties of indirect reports. Note also that the fact that indirect reports involving slurs manifest pervasive, systematic ambiguity between speaker- and attributee-oriented contexts is predicted by expressivism. Indirect quotations typically aim to preserve the truth-conditionally relevant content of what is being reported. Uses of slurs and bare expressives within them enable speakers to both insert their own attitudes or, via implicit quotation, more accurately convey those of their attributees.

what someone said without resorting to explicit quotation, free indirect discourse can be seen as a narrative device to encode what a character thinks and feels without resorting to explicit or implicit 'quotation' of a thought content. Consider the following straightforward discourse:

> Dick was on a mission that demanded extreme caution. He strode into the bar and scanned the crowd warily. Immediately, he spotted his next contact, Henry, in uniform and hunched over a beer. He is a Kraut!, Dick thought. Zbig ambled in moments later, brawny and moustached. Dick watched. Zbig's thick Eastern European accent unnerved the waitress when he barked orders for a vodka. And he, Dick thought, is a Pollack.

and its transposition into free indirect discourse:

> Dick was on a mission that demanded extreme caution. He strode into the bar and scanned the crowd warily. Immediately, he spotted his next contact, Henry, in uniform and hunched over a beer. He was a Kraut! A goddamn Kraut! Zbig ambled in moments later, brawny and moustached. Dick watched. Zbig's thick Eastern European accent unnerved the waitress when he barked orders for a vodka. And he, a Pollack!

Free indirect discourse jettisons the explicit attributions "Dick thought" in the ordinary discourse and shifts the tense of the contained clause ('*is* a Kraut') to that of the narration ('*was* a Kraut!'). Though this leaves open the possibility that the relevant passages could be construed as part of the *narrator's* perspective, the discourse primes the reader to presume Dick's point of view so that we naturally interpret them as if we've been slipped inside Dick's mind.

Free indirect discourse has important advantages over indirect quotation. Most pertinent here: it enables the author to use interjections, vocatives, swearwords, and exclamations—any expressions that cannot be used in subordinate clauses—to flesh out the inner world of the character. Our free indirect discourse narrative could continue:

> Zbig turned and caught his eye. Dick froze. Uh…Oh god. No. Fuck! Fucking fuck!

There is no clear way to otherwise capture so specifically our character's mental state and feelings here, as well as that represented by "A goddamn Kraut!", within an ordinary discourse limited to indirect quotation of contents that are grammatical as subordinate clauses.

Return to (39): "According to Dick, Henry was a Kraut and Zbig was a Pollack". I think it is no coincidence that the sentence that most forcefully demands we interpret it with an attribute-oriented perspective effectively includes part narration 'According to Dick' and a content that is in the same tense as the narration, "Henry *was* a Kraut", not "Henry *is* a Kraut". By evoking the structure of free indirect discourse, it encourages us to treat the 'according to' operator as a device that takes us inside Dick's stance in just the way we shift from narrator's to character's stance in free indirect discourse. The narrator of our toy story is not encoding her own attitudes of contempt, only Dick's.

Exactly how do the slurs and the bare expressives in free indirect discourse convey the character's specific mental state? I won't broach this fascinating (and intimidating) topic here. But what's clear is that free indirect discourse forces us to interpret as if we've been slid inside the character's mind. Any analysis of Dick's mental state regarding what he infers about Henry has to construe it as a thought-feeling complex. One that treats Dick's thought about Henry as "He is German, and worthy of contempt on account of being German!" seems wooden at best.

The last problem Bach advances is that hybrid expressivism entails that anyone who is a Jew is a Kike. It does indeed, but this isn't necessarily a failing of the theory. One can be swept into thinking this consequence is disastrous if one's not careful about how one frames the questions. It might appear innocuous to kick off discussion of our semantic questions by asking what Bach describes as the non-semantic question, "what's the difference between being a Jew and being a Kike?" In my view, this prejudicially misframes the linguistic issues at stake. It is in *many* ways analogous to what has gone wrong when, in investigating the semantic differences between

(41) Is he a homosexual?
(42) Is he a goddam homosexual?

(1) Jake is a Jew.
(33) Jake is a dirty Jew.

we kick off discussion by asking the (so-called) non-semantic question "what is the difference between being a homosexual and being a goddamned homosexual, a Jew and dirty Jew?" This isn't the right kind of question to ask.

True, hybrid expressivism entails that Jews are Kikes.[14] Certainly, we may—should—be uncomfortable with *saying so*, for making such a claim *feels* like an expression of anti-Semitism. Why else would someone say *that*? Yet we're engaged here in a specifically philosophical context, accounting for the linguistic properties of slurs. In this highly circumscribed context, it is worth reminding ourselves that though hybrid expressivism entails that all Jews are Kikes, that amounts to no more than that 'Jew' and 'Kike' have the same extension. Furthermore, comparatively, it hard to find much comfort in loaded descriptivism's determination that "Jews are Kikes" lacks truth value.

---

[14] I offer a full analysis of why a semantics of slurs has this consequence in Jeshion (ms).

References

Bach, K. 2017. "Loaded Words: On the Semantics and Pragmatics of Slurs." Forthcoming in E. Sosa (ed.) *Bad Words*. Oxford: Oxford University Press.

Bell, M. 2013. *Hard Feelings: The Moral Psychology of Contempt*. Oxford: Oxford University Press.

Copp, D. 2001. "Realist Expressivism: A Neglected Option for Moral Realism." *Social Philosophy and Policy* 18: 1–43.

Gutzmann, D. 2013. "Expressives and Beyond: An Introduction to Varieties of Use-Conditional Meaning." In D. Gutzmann and H. Gartner (eds.) *Beyond Expressives: Explorations in Use-Conditional Meaning*. Leiden: Brill.

Hom, C. 2008. "The Semantics of Racial Epithets." *Journal of Philosophy* 105 (8): 416–40.

Hom, C. 2010. "Pejoratives." *Philosophy Compass* 5: 164–85.

Hom, C. and May, R. 2013. "Moral and Semantic Innocence." *Analytic Philosophy* 54: 293–313.

Jeshion, R. 2013a. "Slurs and Stereotypes." *Analytic Philosophy* 54: 314–329.

Jeshion, R. 2013b. "Expressivism and the Offensiveness of Slurs." *Philosophical Perspectives* 27: 231–259.

Jeshion, R. 2016. "Slur Creation, Bigotry Formation: The Power of Expressivism." *Phenomenology and Mind* 11: 130–139.

Jeshion, R. 2017. "Slurs, Dehumanization, and the Expression of Contempt." Forthcoming in E. Sosa (ed.). *Bad Words*. Oxford: Oxford University Press.

Jeshion, R. Ms. "The Truth about Slurs." Manuscript.

Kaplan, D. 2005. "The Meaning of 'Ouch' and 'Oops.'" Unpublished widely circulated manuscript.

Mason, M. 2003. "Contempt as a Moral Attitude." *Ethics* 113: 234–272.

Potts, C. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.

Potts, C. 2007. "The Expressive Dimension." *Theoretical Linguistics* 33 (2): 165–97.

Richard, M. 2008. *When Truth Gives Out*. Oxford: Oxford University Press.

Saka, P. 2007. *How to Think About Meaning*. Berlin: Springer.

Sennet, A. and Copp, D. 2015. "What Kind of Mistake Is It to Use a Slur?" *Philosophical Studies* 172: 1079–1104.

Precis of the Theoretical Part of A Word Which Bears a Sword

NENAD MIŠČEVIĆ
University of Maribor, Slovenia
Central European University, Budapest

Pejoratives are negative terms for alleged social kinds: ethnic, gender, racial, and other. They manage to refer the way kind-terms do, relatively independently of false elements contained in their senses. This proposal, presented in the book, is called the Negative Hybrid Social Kind Term theory, or NHSKT theory, for short. The theory treats the content of pejoratives as unitary, in analogy with unitary thick concepts: both neutral-cum-negative properties (vices) ascribed and negative prescriptions voiced are part of the semantics preferably with some truth-conditional impact, and even the expression of attitudes is part of the semantic potential, although not necessarily the truth conditional one. Pejoratives are thus directly analogue to laudatives, and in matters of reference close to non-evaluative, e.g. superstitious social kind terms (names of zodiacal signs, or terms like "magician"). A pejorative sentence typically expresses more than one proposition and pragmatic context selects the relevant one. Some propositions expressed can be non-offensive and true, other, more typical, are offensive and false. Pejoratives are typically face attacking devices, although they might have other relevant uses. The Negative Hybrid Social Kind Term proposal thus fits quite well with leading theories of (im-)politeness, which can offer a fine account of their typical pragmatics.

Keywords: Pejoratives-slurs, Negative Hybrid Social Kind Term theory of pejoratives, reference, social kind terms.

1. *Introduction*
a) *The main proposal*

This is the precis of Nenad Miščević and Julija Perhat's collection of papers, *A Word which Bears a Sword. Semantics, Pragmatics and Ethics of Pejoratives* (Kruzak, Zagreb 2016), concentrating on the theoreti-

cal part of it; for reasons of space I unfortunately have to leave aside the interesting and challenging contributions of Mirela Fuš, who is basically criticizing my proposal in her "Pejoratives as Social Kinds: Objections to Miščević's Account", Julija Perhat ("Pejoratives and Testimonial Injustice") and Ana Smokrović ("Hermeneutic Injustice and the Constitution of the Subject"), both connecting the use of pejoratives, in particular the gender ones, with testimonial and hermeneutic injustice. Last, but not least, let me mention Martina Blečić writing in more general terms about the connections between pragmatics and ethics ("Slurs: How Pragmatics and Semantics Affect Ethics"). This precis is constructed out of the book material (the long section "The fiery tongue–The Semantics and Pragmatics of Pejoratives" due to Nenad Miščević) re-proposing some of the main thesis of the book.

Pejoratives or slurs, I shall use the two words indiscriminately, are devices for face-attacking, as this term is now standardly used in theories of impoliteness. Jonathan Culpeper, for example, places them on the list of "impoliteness formulae" conventionally associated with offense (2011: 56). Complaints about such offensive communication form a rich tradition, from Saint James Epistle to the present day politically engaged thinkers that Julija Perhat and Ana Smokrović discuss in their respective contributions.

The "Fiery tongue" offers a positive proposal. Pejoratives are negative terms for alleged social kinds: ethnic, gender, racial, and other. They manage to refer the way kind-terms do, relatively independently of false elements contained in their senses. I call the proposal Negative Hybrid Social Kind Term theory, or NHSKT theory, for short. The theory treats the content of pejoratives as unitary, in analogy with unitary thick concepts: both neutral -cum- negative properties (vices) ascribed and negative prescriptions voiced are part of the semantics (preferably with some truth-conditional impact), and even the expression of attitudes is part of the semantic potential, although not necessarily the truth conditional one.

Why believe in unity? First, presumably, the speaker using a P-sentence wants to demean the target on a series of interconnected grounds: X's are bad because of such and such qualities, therefore, we should treat them so-and-so, and therefore I feel contempt for them, and invite you to join in it. Pejorative is not like "ouch", just expressing an inner feeling; inner attitude is grounded in a way of seeing the target, and the way of seeing dictates the attitude. Pejoratives behave in the way one would expect on the basis of interconnectivity of components. When one passes from mild to strong pejoratives, all components change in unison. Similarly, a good translation has to preserve all of the levels: from reference, through specific valuation to expression. Metaphorical (and metonymical) origins of pejoratives also testify to a holistic mapping. Take "Hun", and old British pejorative for Germans. First, we have the vehicle, historical Huns: cruel primitive warriors,

who are very dangerous to potential victims, hated by them, and have to be fought to death by them. Next, the target: Germans, allegedly cruel primitive warriors, who are very dangerous to us, and have to be fought to death by us; of course, we hate them with good reason. If the components were independent, the holism would be unexplainable.

The unity of content also nicely fits the moral phenomenology of negative evaluation. Note that evaluation is not like cheering, supporting my club, and booing the other. The German-hater sees Germans as being in fact bad, not just being guys he is against (as in football). This goes well with wide speech act potential, both illocutionary and perlocutionary, not limited to expressing solidarity with other boo-shouters. The pejoratives can figure in a 2nd person offense, 2nd person expression of solidarity ("I am with you, against these damned Huns!"), 3rd person both offense and solidarity, plus different prescriptions-suggestions that go with it. Only a fixed, semantic meaning can survive this variety, and thrive in it.

What about the truth-value of pejorative sentences, like "Lessing was a Hun"? Each of them express several propositions, some of them true (Lessing was a German), some false (Lessing had such-and-such negative character traits due to his nationality). The proposal is then briefly generalized to other descriptive-evaluative terms, above all to laudatives. Along the way, some well-known puzzles about pejoratives are addressed: the figurative origin of many of them, their occasional positive use by targeted social groups, the role of prohibition in relation to the "bad content", the possible link with cognitive linguistics and more.

b) *The Central Dilemma: empty or literally true*

One can organize the discussion of pejoratives around an important dilemma haunting the theories of pejoratives. Here is one horn of the dilemma:

Pejoratives do refer. Boches are German, period. It's a plus for the theory, since we normally don't think that they are empty. However, the consequence is that the typical basic pejorative sentences are true, since the pejorative does refer, and the sentence ascribes to the target his or her actual belonging to the actual group: Hans is a Boche, The Boches are German, so the truth is secured. Pejorative sentences are simply and literally true, the bad stuff is not truth-conditional, and pejoratives do refer, simply and literally. Call it the veridicality view.

And here is the other, for those who do not like the idea that many typical pejorative sentences are true. The opposite line claims that the reference is empty: there are no Boches, faggots, and so on. The pejorative stands for thick concept, so the negative component is essential to it; since (we know that) no group does satisfy the negative component, the concept is not satisfied, therefore empty.

Each horn is quite unappealing. For the first horn, the veridicality option, part of the trouble is that the assertion of badness is not just

a by-the-way comment, the way it is presented, say by implicature accounts. There, the badness is always part of expressing the standpoint of the speaker, independent from the main topic, the "at issue content" as called by Potts. The leading model for such a reading is offered by general expressive expressions, and sentences like "The fucking dog is again on the couch". Here, the epithet does not primarily characterize the dog, but rather the speaker's attitude to it. And the at-issue content is that the dog is on the couch. But "Boche" and "Čefur" are not like that. The attitude of the speaker is there, but backgrounded. What is in foreground is that the person is bad because German (or former Yugoslav), that is the "issue" in contrast to the typical implicature CI-reading, where the at-issue content is just that the person is German (or former Yugoslav). The expressive dimension is present, but not crucial.

Consider now the non-veridicality side. It is in the clear with the falsity of the P-sentences. The minus is having empty reference for pejoratives; they do not refer simply and literally, they purport to refer, but there is a problem. The dominant ordinary intuition is that pejoratives do have reference. Moreover, how do they offend, if there is no-one to be justifiably offended? The mere clash of intuitions does not solve the problem.

This is then, in my view, the Central Dilemma for the semantic view of pejoratives.

2. *The way out*

Let me point out a way to cope with conflicting intuitions about the truth of pejorative sentences. Take "Lessing was a Boche" (or Hun). The speaker who asserts it shows his knowledge of Lessing's nationality; he cannot be accused of ignorance. On the other hand, we don't want to accept the consequence that, yes, the proposition expressed is true, period. And we want to avoid the specter of disquotation, and the perspective of having to agree that Lessing was a Boche.

Here is the first step to a possible way out: not all propositions expressed by pejorative sentences are false. Some defenders of the implicature view recognize several propositions suggested by a P-sentence. One is the neutral and true sentence (Lessing was a German), other are nasty and problematic. They prefer the neutral one as semantically basic, which I must admit I find counterintuitive. So, I want to borrow from them the general idea that pejorative sentences express a plurality of propositions. Here is the minimum. "*L. was a Boche*" expresses at least 2 propositions:

L. was a German. (the true and decent proposition)
L. was cruel because German. (the false – and indecent – one)

Together with other pluralists, Potts and Bach, I suggest that the context can stress one or the other of the propositions, but in contrast to the first two authors I deny that the true proposition is basic.

We now pass to the main question: if the negative content is part of the meaning, and even of truth-conditional meaning, how can the theoretician avoid the problematic corresponding horn of the Central dilemma, namely that the term is empty, with all its counterintuitive consequences?

An obvious way out is to detach the matter of reference from the matter of the literal truth of claims characterizing the pejorative concept. Causal theories of reference have been suggesting this strategy for various terms and corresponding concepts, and we should turn to them to solve the Central dilemma.

So, what do pejorative meanings or concepts look like? I defend the following proposal: pejoratives are negative (derogatory) social kind-terms, of a hybrid nature. Their referential apparatus involves a causal history of naming plus descriptive senses. The latter have a neutral part (given by a neutral description, German, African, female, gay) and the bad part (primitive, hateful, stupid, etc.), plus perhaps more. It is a social kind-term hypothesis. Surprisingly, it predicts the trouble with the truth-status, in virtue of the neutral/negative contrast: a part of the descriptive sense is neutral, and could (co)ground reference, the other part is negative and introduces the issue of falsity.

What we have encountered until now are at least two layers of meaning or meaning-like dimensions of pejoratives:

First, the minimal descriptive layer, which normally gives the factual information about the target group, and contributes to securing the reference of the term: *African,* gay, and so on, for the corresponding derogatory terms. Second, the negative descriptive-evaluative layer, which ascribes bad properties ("vices") to the members of the target group and often insinuates that they have these properties in virtue of their belonging to the group.

Of course, this is not the end of the story. Many authors, for instance M. Richard, point out that the use of pejoratives often involves a prescriptive suggestion: the target is to be despised (Richard 2008: 15), others would add avoided, or discriminated against, because they exemplify the properties from the negative descriptive content. I am leaving an empty row, since I want to talk more about securing reference later, and will argue for a zero-level of meaning, having to do with the causal-historical link to the group. With these layers we have the minimal material to understand sentences like "L. was a Boche":

L was German.
L belonged to the nation consisting of people, known as Germans, who are cruel because they are members of this nation.

shorter:

L. was cruel because German.

And we know that there is no nation of which the citizens are cruel just in virtue of belonging to it.

THE THREE SEMANTIC LEVELS		
EXAMPLE – "Nigger"		
	LEVEL	CONTENT
SEMANTICS	MINIMAL DESCRIPTIVE	African-American
	NEGATIVE DESCRIPTIVE-EVALUATIVE	primitive, lazy, dangerous
	PRESCRIPTIVE	to be avoided, discriminated against!

The three layers together give the pejorative content a certain "thickness": the word obviously expresses a content that ties together the descriptive and evaluative components, adding the prescriptive dimension connected with the latter. Christopher Hom has rightly insisted on the idea that these contents are "thick concepts" in the technical sense used in discussions in ethics (see Hom 2008). I concur with him on this point, but want to avoid what I see as the weakest point of available semanticist accounts, namely the counterintuitive consequence that pejoratives have empty reference.

Finally, we have the emotional-expressive content: using the pejorative, the speaker expresses his own negative emotional attitude to the target. The dimension of contempt, of placing the target not only in the negative region, but also in the region significantly "below" the (self-assumed) level of the speaker, and thus marking her as "despicable", can partly account for the offensiveness of some pejoratives, and their role in face-attacking verbal acts ("nigger" being the most infamous one). The expression of such offensive attitudes is akin to non-verbal insults, like spitting in someone's face. No wonder that this offensiveness is sometimes described as "ineffability".

3. *Reference and the purely descriptive content*

We first have to address the issue of reference, before passing to the details of the characterization of the bad evaluative content. We have noted that reference of, say "Boche" should be independent of the joint truth-value, i.e. of falsity, of all components of the content of its pejorative meaning; it certainly does refer to Germans, no matter what false ideas about them its use does insinuate. Luckily, there are two elements that are each independently plausible, and that, taken together give the result we need. The first is that the typical referents of pejoratives are social kinds, most often real social kinds, like nationality, gender, age-groups and the like, and, more rarely, assumed kinds; this assumption is plausible independently of the Central dilemma and other semantic considerations.

So, I propose that pejoratives (and in particular slurs) are social kind terms. They refer to moderately clearly identified groups of peo-

ple, or to individuals under the guise of belonging to the group; either socially, or gender-based, or psychologically (e.g. through one's sexual orientation). So, it is social kind terms, plus psychological kind terms that we should look at.

However, reference to kinds often is notoriously generally independent from the speaker's incorrect beliefs about the kind, as the behavior of natural kind terms tends to show; centuries ago, people managed to refer to whales in spite of a lot of false beliefs about them. The reference seems to start with ostension, and tends to follow causal transmission links, not ideas people have about the kind. This gives us exactly what we need. Since we propose a hybrid account of reference, and also hypothesize that the referents are social kinds, real or alleged, the task divides itself into two: first, characterize the relevant social kinds, and second, specify the mechanism of reference.

Candidate social-kinds form a rather heterogeneous bunch; items as mutually different as recession, racism, money, war, permanent resident, prime minister, African-American and German appear on the proposed lists of natural kinds. Note that the first items listed are not analogous to typical biological kinds (like fish, or tiger), the last two are. We shall be interested in kinds like the last two, which classify collections of individuals. For our purposes good examples are ethnicity kinds: Croatian, Finn, and Italian, well known targets of pejoratives. Other examples would be women, gay people, members of some presumed race, perhaps age groups (youth, seniors, etc.) and professions (worker, businessman, journalist).

What about reference to social kinds? If pejorative are negative social kind terms, how do they refer? The preferable option would be that their mechanism of reference is parallel to the mechanism of reference for natural kind and artifact kind terms. There has been an act of baptism, involving some kind of ostension, some characterizing of the target (we need it to avoid the *qua*-problem) and then a chain of transmission up to the present users. But is the parallelism tenable? Our main problem, familiar from the Central dilemma is the one of falsity: a lot of descriptions associated with concepts such as BOCHE are simply false about the intended target(s).

The account proposed is not *ad hoc,* nor specific to pejoratives, or even to evaluative adjectives and nouns in general. It can be applied to the latter, in particular to laudatives, but, more interestingly, to some nouns, like "medicine man" that carry problematic content in their presumed meaning.

Let us start with the easiest case, the laudatives. Take "Aryan" as used by a racist believer in the supremacy of the Caucasian "race". The elements are the same as with pejoratives, only the negative valence is replaced with a positive one. The account can be easily extended to them.

Now pass to a different, non-evaluative sort of problematic terms, like "medicine man" or "Libra" that on the one hand seem to refer, and

on the other appear to have false elements in their senses. We assume that social kind terms like "teacher", "German" and "women" are referentially relatively unproblematic. But what about "medicine man", "theta-rays healer", or close astrologico-psychological kinds: "Scorpio" or "Libra"? They seem referentially problematic in the way reminiscent of our problems with pejoratives. However, we need some background for addressing the issue(s).

First, focus on cases where religious or social beliefs mystify the characteristics of typical and defining activities of certain social groups. Consider the term "medicine man". A relevant original group has been thus designated by original speakers (leaving the issue of translation aside). They were performing activities called by them and their audience "casting spells" and were assumed to have magic powers. The last assumption, is I submit, false. The "casting spells" characterization is ambiguous: first, it can mean pronouncing words and performing gestures that actually do produce results in a super-natural way, second, pronouncing words and performing gestures that are believed to produce the results in such a way by the relevant group of people, including the "medicine men" themselves.

Consider now the sentence, concerning three official "medicine men" of a given tribe, O, Lo and Bo: "O, Lo and Bo are medicine men". Is it true or false? Well, what about magic powers? Presumably, O, Lo and Bo do not have magic powers; so it is literally false since they lack magic powers. But, in the mouth of an anthropologist the sentence probably expresses the proposition that the three men do perform the required activities and are taken to have magic powers. This second proposition is true.

Of course, one can object that "medicine man" is ambiguous between two readings, one that merely indicates a profession and the status that goes with it, call it "medicine man$_p$" and the superstitious, magic related one "medicine man$_m$". When the average tribesperson uses the term, she talks of medicine man$_p$, when the anthropologist uses the term, she refers to medicine man$_p$. This is a legitimate understanding, but it leaves out the fact that both talk about the same people, that the anthropologist can try to persuade the tribesperson that these people have no magic powers, and so on. "Medicine man" is not ambiguous in the way in which "bank" is.

Now, take another problematic group, the names of astrological signs, e.g. "Scorpio". The name presumably refers to persons born between October 23 and November 21; it has been transmitted for some thousands of years to the present times. On the other hand, it is also used to refer to people who presumably have such-and-such "Scorpionic" character traits in virtue of being born in the given period of time. Here is a description taken from the web site:

> Scorpio is the eighth sign of the zodiac, and that shouldn't be taken lightly—nor should Scorpios! Those born under this sign are dead serious in their mission to learn about others. There's no fluff or chatter for Scorpios, either;

these folks will zero in on the essential questions, gleaning the secrets that lie within. Scorpios concern themselves with beginnings and endings, and are unafraid of either; they also travel in a world that is black and white and has little use for gray. The curiosity of Scorpios is immeasurable, which may be why they are such adept investigators. These folks love to probe and know how to get to the bottom of things. The fact that they have a keen sense of intuition certainly helps. (http://www.astrology.com/scorpio-sun-sign-zodiac-signs/2-d-d-66949)

I propose that in the context of astrology "Scorpio" is a hybrid name for a presumed, but highly problematic psychological kind, whose reference is determined both causally and descriptively, whereby the descriptive component has two sub-components: the unproblematic, time interval component, and the problematic, superstitious character describing component.

Consider now the sentence: "Nenad is a Scorpio"; is it true or false? Suppose it expresses the neutral propositions:

Nenad is a Scorpio, he was born on November the first.

It is true then. But here is the non-neutral, superstitious reading:

Nenad is a Scorpio, (as regards his character.)

And this one is false.

Interestingly, problems with reference and the plurality of propositions have nothing to do with evaluative elements. Also, the superstitious material is not a comment on the descriptive material, the way in which negative material is seen by CI-theorists as a comment on factual material in the case of pejoratives.

The analogy with such hybrid terms reinforces the main point of my proposal: pejoratives, say "N", are negative (derogatory) social kind terms, with a hybrid nature. Their reference is partly determined by a causal chain: the target group G has been called by somebody "N", the name has been transmitted to the present users, and it refers to the group G and its members. Their descriptive senses have neutral material (given by a neutral description ("German", "female", "gay"), and bad material (primitive, hateful, stupid, etc.) plus more. I have called the proposal the negative hybrid social kind terms hypothesis (NHSKT hypothesis).

4. *The negative content*

On the negative side we have several layers. The first is evaluative, but most often it contains some descriptive or semi-descriptive component: Boches are bad, for specific reasons, namely because of their cruelty and the like. I shall call this mixed layer "descriptive-evaluative", in contrast to a purely descriptive characterization, like "being a German". Next, there are prescriptive and expressive layers, naturally connected to the negative evaluation.

The content at the descriptive-evaluative layer points to bad properties and ascribes them to the members of the target group, normally

with the insinuation that they have these properties in virtue of their belonging to the group. In characterizing the layer, I first consider the properties themselves, both from a more socio-psychological viewpoint, and from anormative philosophical one. Then I turn to the meta-ethical characterization of the content having to do with its "thickness".

So, what is being imputed? Let us focus upon full blooded pejoratives, involving rich negative material. I shall very briefly offer two characterizations. One line one can take in characterizing the negative-descriptive (not prescriptive) content is to liken it to stereotypes.

The prototype associated with full-blooded slur is normally a very negative stereotype. Here are our two examples:

EXAMPLE	"Boche"	"curr"
MINIMAL DESCRIPTIVE	German	dog
NEGATIVE DESCRIPTIVE-EVALUATIVE	cruel and dangerous because German	dangerous, of low origin, etc.

As mentioned, I take the second row to exemplify a dimension of meaning of pejoratives, i.e. the negative descriptive-evaluative one. I assume gradation in badness since some pejoratives are more devaluing than others, for example in English: the negative value implied by "minx" is not as dramatic as the one belonging to "whore" (Hughes 2006: 163). We can have a plurality of closely related dimensions, say typical negative properties plus their degree of badness.

The second way to characterize the negative descriptive-evaluative content is to connect it to virtue ethics. Negative stereotypical traits will be then classified as vices. Vices are often characterized as qualities that both attain bad ends or effects, and involve bad motives. For an illustration, along more traditional lines, consider the pejorative use of "pagan" in Pope's injunction from St. Stephen's day 2014: "Don't live like pagans, live like Christians!" Living like Christian includes practicing virtues like generosity, sexual moderation and the like. "Pagans" are persons that have lived with vices of greed and lust, presumably *because* they have not believed in the true god (or have not believed in god at all).

Let me now pass to thick concepts. Concepts uniting neutral descriptive and evaluative components have been traditionally classified as "thick concepts". Thick concepts play important roles in various domains of evaluation. When evaluating a policy in prudential terms we sometimes describe it as wasteful, stating that it wastes resources and implying that it is therefore less than adequate. Decisions are sometimes criticized as rash, people as being greedy. On the epistemic side, a proposal might be praised as thoughtful, and an idea as deep. On the esthetic side, thick concepts are the building blocks of art criticism; think of ones like ELEGANT, KITCH, or TOUCHING.[1] Thick

---

[1] I will adopt the convention of writing concept terms in capitals.

concept-words are often likened to serious pejoratives ("Kraut", "faggot"), another topic of quite intense research, although some authors deny similarity. However, most of the work done on thick concepts has been dedicated to moral ones, depicting virtues and vices, like COURAGE(OUS), CRUEL(TY), LEWD, NOBLE.

The original story of thick concepts, as told for instance by B. Williams (1985), was that they carry the (moral-) evaluative attitude on their sleeves, plus that the attitude is fixed within a very narrow range: courage is admirable, period. Change the attitude, and the concept is gone.

The minimal form of thickness involves the unity between the descriptive and the evaluative. And indeed, the properties associated with pejoratives are evaluatively rich properties posing as objective properties of the target. Even expressivists like Richard agree that pejoratives present their targets under a negative guise, and that the negative guise is not merely a general negative characteristic (the target is bad) but a rich more specific characterization (primitive, dangerous …). Note that even this minimal form of thickness creates problems for the separationists who propose an expressivist reading of the evaluative component.

We have noted the analogy between pejorative concepts and the usual ethical thick concepts. But, there is an important difference between central, paradigmatic thick concepts and pejorative, slurish if you like the term, contents or concepts. Paradigmatic thick concepts are general, and centrally related to adjectives; they are the contents of corresponding adjectives, not nouns. They just tell us about a presumed property, and it is an open question which kinds of entities carry the properties. The non-empty domain is not guaranteed: perhaps, there are lewd behaviors, lewd shows and the like, perhaps not. Things stand differently with generic pejoratives-slurs. They primarily target some *given, independently identified* group, and their content is tied to nouns rather than to adjectives. The German hater starts with referring to Germans, and then goes on to ascribe cruelty to them, the gay hater starts with referring to gays, and then proceeds to suggest their presumed negative properties. This is why reference and reference-determining material is independent from the negative (evaluative, prescriptive etc.) features, and why pejoratives are (unfortunately) not empty.

We now pass to the further layer that naturally goes with evaluative thickness, namely the prescriptive one. Value and prescription normally go together; this is the first thing one learns in normative ethics. With pejoratives it is the negative valuation that counts. Badness intrinsically repels the agent who understands it, and so on. To put it in nowadays usual form: at the least, the badness of X gives a prima facie reason to avoid (doing, encountering, having to do with) X.

We need a very modest form of this claim. First, we can rest satisfied with the phenomenological dimension: if our racist finds (experiences)

some qualities sufficiently bad, this will give him, at least from his perspective, a prima facie reason to act in the way of avoiding, downgrading etc. items (things and people) whom he experiences as having these qualities. And, if he is consistent, he will be motivated to do it.

This brings us to the topic of thickness. The link to prescription is very strong in (the standard picture of) thick concepts: they essentially engage in "action guiding". Now, if we accept the minimal thickness and add this link to the prescriptive component (and to queerness) we obtain a richer form of thickness. It encompasses motivation and prescription. What would be the message for the semantics of pejoratives? The connection between negative value and corresponding prescription hold as well in the case of pejoratives, and points to the unity of pejorative meaning. Many authors point out that the use of pejoratives often involves a prescriptive suggestion: the target is to be despised, others would add avoided, or discriminated against, because she exemplifies the properties from the negative descriptive content. "Fags will burn in Hell" is a well-known variant of such prescriptivism, directed to the future and eternal suffering of gays. I will leave matters at this, but the interested reader might wish to consult Perhat's chapter for further material.

Here is then the summary in the form of a table:

THE FINAL PROPOSAL: MAXIMAL SEMANTICIST EXAMPLE – "Nigger"		
	LEVEL	CONTENT
SEMANTIC	CAUSAL-HISTORICAL	someone called them thus
	MINIMAL DESCRIPTIVE	African-American
	NEGATIVE DESCRIPTIVE-EVALUATIVE	primitive, lazy, dangerous
	PRESCRIPTIVE	to be avoided, discriminated against!
PRAGMATIC	EXPRESSIVE	Yuck! and more

What is then the content expressed by a pejorative sentence? A typical content such sentences suggest is a plurality of propositions, in which the factual and the bad-material propositions are on equal footing, both of them are truth-apt and equally well expressed by the pejorative sentence. The interest in context picks out the relevant proposition, and is responsible for treating the sentence sometimes as true (when the bad material is not in focus), sometimes as false (when the bad material is in focus).

Where do we go on from here? I have already mentioned the socio-pragmatic framework of impoliteness research developed mostly by linguists and anthropologists (Leech 2014, Brown and Levinson 1987, Culpeper 2011). I believe there is theoretical unity and interconnected-

ness that goes all the way from the semantics of pejoratives, through their pragma-linguistics (speech act theory), to socio-pragmatics, impoliteness theory and rhetoric. A natural further step would be to try to unify the proposed semantic explanation(s) with their possible pragmatic counterparts (which I try to do briefly in the last chapter of my contribution ("Using the Verbal Poison: Pejoratives and Impoliteness-Rudeness"). Another obvious direction would be to extend the semantics (and pragmatics) of pejoratives to their symmetrically looking counterparts, laudatives, which would enhance the theoretical unity of the account(s).

So much for the main theoretical proposal. For interesting developments, criticism and ethical and political applications see the papers by other collaborators in the book.

References

Bach, K. 1999. "The Myth of Conventional Implicature." *Linguistics and Philosophy* 22 (4): 327–366.

Brown, P. and S. C., Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press (first published 1978).

Culpeper, J. 2011. *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.

Hughes, G. 2006. *An encyclopedia of swearing: The social history of oaths, profanity, foul language, and ethnic slurs in the English-speaking world*. Armonk: M.E. Sharpe.

Miščević, N. and J. Perhat. 2016. *A Word which Bears a Sword. Semantics, Pragmatics and Ethics of Pejoratives*. Zagreb: Kruzak.

Richard, M. 2008. *When Truth Gives Out*. Cambridge: Cambridge University Press.

Potts, C. 2005. *The Logic of Conversational Implicatures*. Oxford: Oxford University Press.

Hom, C. 2008. "The Semantics of Racial Epithets." *The Journal of Philosophy* 105: 416–40.

Leech, G. 2014. *The Pragmatics of Politeness*. Oxford: Oxford University Press.

Williams, B. 2013. *Ethics and the limits of philosophy*. Cambridge: Harvard University Press (first published 1985).

Pejoratives and Testimonial Injustice: Precis

JULIJA PERHAT
Faculty of Humanities and Social Sciences,
University of Rijeka, Rijeka, Croatia

Testimonial injustice is a hot topic in social epistemology. My own work is concerned with pejoratives (in particular, gender pejoratives for women), so in this paper I wish to connect them with such injustice. So, my present topic is testimonial injustice perpetrated by the serious use of pejoratives, in particular, gender pejoratives. It combines two strands: on the one hand, the work on testimonial injustice; and here I shall rely on Miranda Fricker's work, and on the other hand, my own central area of interest, (gender) pejoratives.

Keywords: Gender pejoratives, testimonial injustice, linguistic injustice, stereotypes.

1. Introduction

In this work I try to connect testimonial injustice to gender pejoratives. In order to that I first briefly explain what pejoratives are. Then I move on to explaining Fricker's view on prejudice and stereotypes. I would like to show how Fricker's idea of collective social imagination is indeed very plausible and how stereotypes and prejudice are a normal part of it. These stereotypes and prejudice typically activate themselves when one tries to assess someone's credibility or trustworthiness and the activation of stereotypes and prejudice happens even if one is not aware of it. Now, I think that language is an important part of socialization and can increase the testimonial injustice. Since stereotypes and prejudice have the main role (as Fricker showed in her book Epistemic Injustice) in the process of attributing credibility to the speaker, or in other words, sexist prejudice decrease the credibility assigned to women, I try to show how the use of pejoratives can develop and sustain stereotypes and prejudice and thus influence our perception of the speaker. Furthermore, not only can pejoratives have an influence on our percep-

tion of the speaker but their permanent and systematic use has an influence on the systematic underestimation of credibility of certain groups of people (and consequently on their self-esteem and other intellectual virtues). In this precis I am particularly interested in how pejoratives influence the underestimation of credibility of women and I think that the use of pejoratives can indirectly and directly increase the testimonial injustice in the second step and how they perpetuate the testimonial injustice done to women. So, the question is how should we cope with such phenomena in our society? In the last part I try to give a possible answer to that question by presenting and pointing out possible virtues that we should have in order to avoid the use of pejoratives and I also try to give an answer to the question about the testimonial sensibility that we should have in order to avoid the use of pejoratives. Fricker claims that we should develop testimonial sensibility and I point out virtues that can help in that process. The idea is that one possible virtue that we may want to nurture is the love of knowledge so that we could avoid the non-culpable mistakes. Another virtue that we should have in order to avoid pejoratives as hearers of such language is the virtue of autonomy, and, finally, it is evident that it would be virtuous not to use pejoratives and that users of pejoratives obviously have testimonial insensibility instead of testimonial sensibility as Fricker argues in her book Epistemic Injustice.

2. *Pejoratives*

There is a wide array of pejorative words, but here I am concerned with those that are gender related, namely sexist pejoratives for women. Some of the examples of sexist pejoratives are English pejoratives: *bitch*, *cunt*, *whore* and *witch*; and their Croatian synonyms would be: *kuja*, *pička*, *kurva* and *vještica*.

So, let me briefly give a semantic analysis of the word *whore*. There are three meanings of the word *whore*.

The first one is the literal meaning where the word *whore* means prostitute. A prostitute sells her body for money which is considered to be bad and immoral.

The second meaning refers to a promiscuous woman who sleeps with a lot of men, but the reference to money is no longer a part of the meaning (she, unlike a prostitute, does not "sell" her body), so we can say that this second meaning of the word *whore* is a half dead metaphor. But, the qualities that did stay represented within the very meaning (sense) of the word *whore* are such that the woman who the speaker is referring to by the term in question is bad and immoral because she sleeps with a lot of men. So, something negative and devaluating is rooted in the very meaning of the word *whore*. When someone calls a woman a *whore* what is rooted in the meaning of the word is that she is bad, immoral, dirty… So, when someone uses the word *whore* to refer to a woman, in its presumed extension it means that she will sleep with almost anyone (promiscuous behavior), that she is easy, etc.

Used in its third meaning the word *whore* can apply to "anyone who sells out their principles" (Hughes 2006: 493). So, it does not need to refer to women, it can also refer to politicians, for example. This third meaning of the word *whore* has no sexual reference. However, it does mean that the target is corrupt or immoral because of selling their principles.

Recently, a new meaning of the word *whore* has arisen. It can be classified as the fourth meaning of the word (*reference of which is* the *Urban dictionary*). It refers to a person that is doing something excessively and repeatedly which is very annoying. The link to the literal meaning of prostitute is obviously lost (as it is in the third meaning), but we can find the link with promiscuity (even though the sexual component is lost) because promiscuity can signify repetition which is considered to be annoying and bad.

The most interesting meaning of the word *whore* for the discussion here is the second one. In the corresponding case what the speaker intends to do is to degrade the target (a woman whom he refers to as a *whore*).

Now I want to give a background on socialization and stereotypes, and also, a short overview of empirical material on stereotypes which I will then connect with Fricker's theorizing, and, also, address the issue of defining stereotypes.

3. *Miranda Fricker's outline of testimonial injustice*

I am relying on Fricker's idea of testimonial injustice which she claims to be a normal part of the discourse, unfortunately. Here, we need some background, so I briefly summarize her main points that I find useful for my project. Before Fricker goes on to explain what exactly testimonial injustice is, she turns to the idea of identity power which impacts our discursive relations.

3.1. *Power and testimonial exchange*

Fricker explains that there is "at least one form of social power which requires not only practical social co-ordination but also an *imaginative* social co-ordination" (Fricker 2007: 14). This is where Fricker presents us with an idea of identity power:

> There can be operations of social power which are dependent upon agents having shared conceptions of social identity – conceptions alive in the collective social imagination that govern, for instance, what it is or means to be a woman or a man, or what it means to be gay or straight, young or old, and so on (Fricker 2007: 14).

Fricker gives an example of gender acting as one arena of identity power and stresses how an *active* use of gender power can be when a man uses his identity as a man to influence woman's actions. The example that Fricker presents us with is the case where a man silences

a woman by emphasizing that she cannot possibly be right about her suspicions about the possible murderer because, as he explains, all of her suspicions are only based on female intuition as opposed to facts (Fricker 2007). Also, identity power can be structural or agential.

But, how does identity power influence testimonial injustice. Here comes the thought that will be central in the sequel. Fricker writes:

> I shall argue that identity power is an integral part of the mechanism of testimonial exchange, because of the need for hearers to use social stereotypes as heuristic in their spontaneous assessment of their interlocutor's credibility. This use of stereotypes may be entirely proper, or it may be misleading, depending on the stereotype. Notably, if the stereotype embodies a prejudice that works against the speaker, then two things follow: there is an epistemic dysfunction in the exchange – the hearer makes an unduly deflated judgment of the speaker's credibility, perhaps missing out on knowledge as a result; and the hearer does something ethically bad—the speaker is wrongfully undermined in her capacity as a knower (Fricker 2007: 16–17).

Now, prejudice can result either in *credibility excess* (when "a speaker is given more credibility than she otherwise would have") or in *credibility deficit* (when "a speaker receives less credibility than she otherwise would have") (Fricker 2007). However, what will be of most interest here is the credibility deficit which can lead to testimonial injustice. In the cases of testimonial injustice the ethical poison "must derive from some ethical poison in the judgment of the hearer... The proposal I am heading for is that the ethical poison in question is that of *prejudice*" (Fricker 2007: 22).

This brings us to the possible link with pejoratives, namely the importance of the stereotype. I want to point out the connection between gender pejoratives and the way they can indirectly increase the testimonial injustice done to women and that is already part of the imaginative social co-ordination.

3.2 *Prejudice and stereotypes*

First, how should we define stereotypes? Fricker is wisely not prejudging the badness of pejoratives: so, since she is using the word stereotypes neutrally, she will need a broader definition. Indeed, she defines stereotypes as "widely held associations between a given social group and one or more attributes" (Fricker 2007: 30). When defined like that "stereotyping entails a cognitive commitment to some empirical generalization about a given social group ('women are intuitive')" (Fricker 2007: 31) and that generalization can be more or less strong.

She goes on to say that we can suppose that an identity prejudice is at work in the stereotype. In that case we have to understand that a stereotype can be just a non-culpable mistake (an 'honest mistake'). Arplay (2003) gives a good example of a boy who has a belief that women are not capable of abstract thinking, at least not as men are. Since he lives in a community where all the evidence he could have gathered

suggests that women are indeed not capable of such thinking, we can say that he made an honest mistake. If the boy was to come across some counter-evidence and he does not change his belief than we can say that he does something ethically and epistemically bad (Fricker 2007).

In these cases there exists a negative identity prejudice which has an ethically bad motivation behind it, so the identity prejudice that Fricker is focused on and that I will also be examining are "prejudices with a negative valence held against people *qua* social type" (Fricker 2007: 35).

Therefore, Fricker gives a definition of what negative identity-prejudicial stereotype is: "A widely held disparaging association between a social group and one or more attributes, where this association embodies a generalization that displays some (typically, epistemically culpable) resistance to counter-evidence owing to an ethically bad affective investment" (Fricker 2007: 35).

Now, Fricker goes on to say that identity prejudice "distorts the hearer's perception of the speaker" (Fricker 2007: 36). To clarify exactly how that can happen Fricker takes the explanation of the idea of a stereotype proposed by Walter Lippmann (1965). He "described social stereotypes as 'pictures in our head'" (Fricker 2007: 37). These can linger on in our psychology and affect the hearer's patters of judgment even when our belief system in not in accordance with this. An example given by Fricker on this point is of a feminist who doesn't take a word of her female colleagues seriously.

Fricker takes this (the stealth mode of social stereotypes which persist in our psychology despite ourselves) to be evidence or at least support the idea that testimonial injustice happens all the time. She agrees with Shklar (1990) that injustice is a normal social baseline, and she thinks that also testimonial injustice is a normal part of the discourse. She also emphasizes the wrong that is done to someone when treated in this way (an ethical wrong that can be damaging) which is still viewed as something trivial (Fricker 2007).

Fricker then moves on to further develop this point. She claims that testimonial injustice can do little or no harm but that it can also be seriously harmful when it is systematic. Fricker recognizes epistemic harm where "knowledge that would be passed on to a hearer is not received" (Fricker 2007: 43). However, Fricker is more concerned with the immediate wrong that the hearer does to the speaker. She emphasizes that the ability to give knowledge to others is significant for human beings. So, when someone suffers a testimonial injustice they are not only degraded as knowers but also as humans. Considering this aspect of harm, she concludes: "The harm will take different forms, but they are both cases of identity-prejudicial exclusion from the community of epistemic trust, and so they both belong to the same category of injustice" (Fricker 2007: 45–46).

Fricker also discusses the secondary aspect of harm where she explores two categories; practical and epistemic dimension of harm. To explain what practical dimension of harm would be she presents an example of a testimonial injustice in a courtroom where one can be found guilty. The second category is again that of an epistemic harm where:

> The recipient of a one-off testimonial injustice may lose confidence in his belief, or in his judgment for it, so that he ceases to satisfy the conditions for knowledge; or, alternatively, someone with a background experience of persistent testimonial injustice may lose confidence in her general intellectual abilities to such an extent that she is genuinely hindered in her educational or other intellectual development (Fricker 2007: 47–48).

All of this can have an impact on the formation of our identity if, of course, the testimonial injustice is persistent and systematic (we can imagine a woman interested in politics but living in a society that doesn't allow women to vote; in such a case, she is to lose an essential part of herself by not being a part of a trustful conversation). One (in our example a woman) may also come to resemble the stereotype working against her (she may internalize the stereotype and start to believe that she is indeed inferior and act according to it).

I would like to focus on the harm done to women in our Western liberal society thus emphasizing that our society is indeed still patriarchal, even if we are not so keen to admit it. And there is a connection between the language we use to speak about women and the way language can indirectly (and sometimes directly) increase the testimonial injustice done to women. Also, I would like to focus on importance of virtues that we should have in order to avoid the use of pejoratives and try to answer the question what is the sensibility that we should train in order to avoid them.

4. *Pejoratives and testimonial injustice*

By calling a woman a *whore* one undermines her dignity by not treating her as an equal member of the society (remember the brief analysis that is offered in section 2). When using such sexist speech the goal of the speaker is to degrade. So, the (intended) perlocutionary effect is to degrade the target and to treat her as less valuable than other members of the society. It is worth noting that it does not matter for the woman if she is physically present at the time of the utterance. The degrading of the target happens even if the target is not present and even if there is no face-to-face confrontation. The term *whore* is equally offensive for a woman whether she is present at the time of the utterance or not. Also, when the speaker uses such terms his intended perlocutionary effect is for the hearers to agree with him, too. It is, of course, clear that the use of pejoratives is not itself a testimonial injustice (there may be some similar consequences such as undermining a woman's self-esteem, which is also the result of systematic skepticism towards women's credibility, but it isn't a testimonial injustice in

itself). Now, the question is how does using pejoratives contribute to testimonial injustice happening to women?[1]

We can take an example of the speaker using pejoratives (the word *whore*, for example) and being in a position of power[2] in which case we can only imagine the scope of the harm being done. Suppose that an executive director of a certain company is talking to his young employees and that he refers to his (and their) female colleagues as *whores* or *bitches*. As I already explained the (intended) perlocutionary effect is that the hearers agree with him. Since the speaker is in a position of power and if the hearers are already mentally contaminated, it is plausible to assume that the pejorative can increase the already existent identity-prejudicial stereotype (which is, as Fricker explained, present in the collective social imagination). It can also be the case (a worst case scenario) that the speaker receives credibility excess in which case the perlocutionary effect on the speakers to agree with the hearer is even greater. But, in any case pejoratives can, as we have seen, increase the already existent identity-prejudicial stereotype thus indirectly increasing the testimonial injustice done to women. After all, who would trust a *whore*?[3] There are cases where using a pejorative can even directly increase testimonial injustice, for example, in cases when uttering: "Shut up, you bitch". Also, to add to Fricker's claim where she described identity-prejudical stereotype as an association

[1] I am here exploring the connection between pejoratives and testimonial injustice. However, it seems that pejoratives can inflict an even greater epistemic harm. When using hate speech (for example, calling someone a whore) the person who is the target of such speech can internalize all the bad things that are meant/implied by the expression in question and thus hate speech may influence the very character of a person (one referred to as a Nigger can indeed start acting lazy, stupid and so on). This very interesting effect of hate speech is something to explore in the future due to the limitation of this paper.

[2] By the term "position of power" I mean both the identity power identified by Fricker and the economical power excersized through the person being an executive of a certain company.

[3] Now, let's imagine a speaker who calls his friend *a whore*, but who is an epistemically disciplined person and he would never use a person's being *a whore* to determine her testimonial worth. However, the people who hear that comment do take her less seriously. It seems that the utterer is not guilty of testimonial injustice, so the question remains: who is guilty of what? (This comment was made by Johanna Schnurr at a conference in Dubrovnik 2014).

So, in this particular case we can speculate that the speaker is not directly guilty of testimonial injustice since, as we have established, he does not undermine the person's testimonial worth, but he certainly is guilty of *indirectly* increasing testimonial injustice since he is using a pejorative in a certain context. This is more of a consequentialist view where the person is contributing in perpetuating the society where certain groups of people can be degraded by using such pejoratives, and consequently indirectly increasing the testimonial injustice because of using such a pejorative to describe a person and he /she should be aware of its effect on the hearers. The only case where the speaker cannot be considered culpable is a case already described by Fricker where a speaker only has evidence that support his prejudicial belief.

which "embodies a generalization that displays some (typically, epistemically culpable) resistance to counter-evidence owing to an ethically bad affective investment" (Fricker 2007: 35), it seems that pejoratives create a context where one is prone to resist any counter-evidence that may occur.

Now, let's note here that there are certain pejoratives which can increase credibility[4] but in this paper I am interested only in sexist pejoratives for women which, I think, usually lower credibility. So, when calling someone a *bitch* or a *whore*, you actually want to degrade (remember that we are concerned with the literal, serious use of pejoratives as opposed to appropriated uses or just mentions—quotes) and if you succeed then the person would not be taken seriously in the future, so you would actually lower her credibility by degrading her. It seems fair to assume that these kinds of pejoratives would attack sincerity more than they would attack knowledge. For example, you could trust a person (who was referred to as a *whore*) when she says that it is raining outside. But, you wouldn't trust her when she would claim to be in love. Although, in my opinion, it also seems fair to assume that calling someone a *whore* or referring to them by some other (gender) pejorative would actually lower their knowledge-status in the sense that they wouldn't be taken seriously in, for example, the workplace (if you think someone is a *whore*, this will probably screen off any considerations of qualities such as being knowledgeable or an expert, and also raise doubts that she got a good status due to her expertise at work suggesting that she got where she is by other, less respectable means).

Another point that I would like to draw is the fact that, when talking about sexist pejoratives for women, the important thing to note is that they go with sexist stereotypes about women. These kinds of pejoratives are interesting because they carry rich content (the negative valence is rooted in the meaning), which means that one word can carry a lot of power and using them to degrade another person is certainly more appealing to most people than using another form of communication (for example than just saying "She is promiscuous and that is bad and immoral"). Because of this my opinion is that we can talk about *linguistic injustice* which can happen when somebody uses a pejorative to refer to (a) wom(a)en. So, when calling someone a *whore*, what you mean is that the person is promiscuous and that is bad, she has a loose moral, she's dirty, and you probably shouldn't get romantically involved with her. In general, for a pejorative to inflict linguistic injustice (and to degrade) it would have to carry a stereotype in its content which, I think, sexist pejoratives do.

Now, the next question to be asked is how to cope with such phenomena? Above all, if the conclusion is that the serious use of pejoratives should be avoided, how should this be done? In previous sections

---

[4] Pejoratives such as *nerd* or *swat*, which was pointed out by T. Williamson at The Linguistics and Philosophy conference in Dubrovnik 2014.

I have argued that pejoratives sustain an unjust system by indirectly increasing testimonial injustice. Now, one way of dealing with pejoratives is to have certain virtues in order to avoid the use of pejoratives; this might help to answer the question about the testimonial sensibility that we should develop in order to avoid pejoratives.

It is important to note that one can use pejoratives and make a non-culpable mistake or one can use them merely as provocation or sarcasm. One can make a non-culpable mistake and a good example is given by Arpaly (in Fricker 2007: 33) where she describes a boy who doesn't have any access to knowledge and therefore it is understandable if he thinks that women are somehow inferior to men. But that is why it is important to have the virtue of love of knowledge and as Roberts and Wood explain: "The virtue of love of knowledge packages a desire for knowledge, along with the sense of the relative importance of truths, and thus which truths merit pursuit. The social side of the love of knowledge includes a willingness and ability to convey relevant truths to others" (Roberts and Wood 2007: 165). So, non-culpable mistakes can happen but if we have a virtue of love of knowledge then as our knowledge grows, the non-culpable mistakes will be fewer.

The worst kind of use of pejoratives would be to use them in order to spread and deepen the injustice (as I have emphasized before, this use of the pejorative and this range of a typical pejorative is of interest in this paper). In order to avoid this we, *as hearers* of such speech, should have the virtue of autonomy and intellectual autonomy disposes us to be appropriately dependent on others' intellectual guidance and achievements (Roberts and Woods 2007: 267). So, in order to reject the use of pejoratives in a hate-filled environment what we should have as a virtue is, I think, intellectual autonomy where we rely on ourselves. In that case we would not just agree with what the speaker who is using sexist speech is saying, so his intent (that we agree with him) would fail.

5. *Conclusion*

I tried to connect the idea of testimonial injustice with the theory of pejoratives. Here is my argument in a nutshell: one way of forming stereotypes is through language because language is an important part of socialization. If we use language to spread hatred (which is the case in using pejorative for degrading the target) then language becomes a harmful weapon. Therefore, pejoratives can indirectly (and directly) increase the testimonial injustice. I have also argued in favor of cultivating some virtues in order to avoid the use of pejoratives, but some things still have to be said about the users of pejoratives.

Thus, the question left unanswered is the question about the user(s) of pejoratives. We can say that the one that uses pejoratives did not engage in self-critical reflection and Fricker also notes that for a hearer to identify the impact of identity power in their credibility judgment, the hearers must also be alert to the impact the speaker's (and their own)

social identity may have on their credibility judgment. However as Fricker (2007) argues, it is not only the user of pejoratives who should engage in self-critical reflection, but also the hearers who should constantly question which prejudice may interfere in the discourse thus avoiding the lack of credibility they assign to the speaker. As noted earlier we all have unconscious processes and that is why this virtue is probably the most helpful one in assessing our own stereotypes and prejudice (even if we are not fully aware of all of our prejudice I think that this virtue can help develop our testimonial sensitivity[5]). Once we realize that we may have certain prejudice we can act in order to avoid them. Also, this virtue decreases the amount of non-culpable mistakes because it makes us constantly question possible prejudice that may influence the assessment of the credibility of the speaker. If one fails to engage in critical self-reflection then one also may fail to recognize the prejudice that is contaminating his belief system (Fricker 2007). As I have already pointed out, this virtue may apply to users of pejoratives because they should also question their belief system and, of course, change it if presented with counter-evidence. I think that the users of pejoratives (with the intent to degrade) do not possess the respect of the autonomy of others. Consider Kant's second formulation of the categorical imperative which states that we should never treat humanity in another person merely as a means to an end. If we do so, we do not respect that person and we violate their autonomy because a person who is treated merely as a means to an end, instead of an end in themselves, cannot be autonomous. The user of pejoratives violates all of the above. It is clear that the user of pejoratives has not trained his/her testimonial sensibility (and the training of our testimonial sensibility is necessary if we were to comprise virtues needed to avoid pejoratives). Quite the opposite, the speaker has testimonial insensibility. It would be virtuous not to use pejoratives (if they are used to degrade).

References

Arplay, N. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press.

Fricker, M. 2007. *Epistemic injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Hughes, G. 2006. *An encyclopedia of swearing: The social history of oaths, profanity, foul language, and ethnic slurs in the English-speaking world*. Armonk: M.E. Sharpe.

Roberts, R. C. and Wood, W. J. 2007. *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford: Oxford University Press.

Shklar, J. 1990. *The Faces of Injustice*. New Haven and London: Yale University Press.

[5] The idea of testimonial sensibility as Fricker explains it is as follows: "Our idea of testimonial sensibility is an idea of a spontaneous critical sensitivity that is permanently in training and continuously adapting according to individual and collective experience" (Fricker 2007: 84).

Social Identity, Indexicality, and the Appropriation of Slurs

KATHERINE RITCHIE
City College of New York, CUNY, New York, USA

Slurs are expressions that can be used to demean and dehumanize targets based on their membership in racial, ethnic, religious, gender, or sexual orientation groups. Almost all treatments of slurs posit that they have derogatory content of some sort. Such views—which I call content-based—must explain why in cases of appropriation slurs fail to express their standard derogatory contents. A popular strategy is to take appropriated slurs to be ambiguous; they have both a derogatory content and a positive appropriated content. However, if appropriated slurs are ambiguous, why can only members in the target group use them to express a non-offensive/positive meaning? Here, I develop and motivate an answer that could be adopted by any content-based theorist. I argue that appropriated contents of slurs include a plural first-person pronoun. I show how the semantics of pronouns like 'we' can be put to use to explain why only some can use a slur to express its appropriated content. Moreover, I argue that the picture I develop is motivated by the process of appropriation and helps to explain how it achieves its aims of promoting group solidarity and positive group identity.

Keywords: Slurs, appropriation, reclamation, indexicals, social groups.

Slurs are expressions that can be used to demean and dehumanize targets based on their membership in social groups based on, e.g., race,

ethnicity, religion, gender, or sexual orientation.[1] Recently there has been increased interest in the semantics and pragmatics of slurs in philosophy of language and linguistics. While accounts vary significantly, almost all theorists agree that slurs have derogatory content of some sort or other.[2] Various theories have been developed to address how derogatory, demeaning, or other negative content is encoded and expressed. For instance it has been argued to be part of truth conditional content, presuppositional content, expressive content, and conventionally implicated content. I'll call all of these views and their variants content-based views.

On content-based views while slurs generally express something derogatory, there are cases of (re)appropriation or reclamation in which groups that are targeted by a slur reclaim it for positive in-group usage.[3] Content-based theorists often argue that appropriation involves meaning change, in particular, it involves an expression becoming ambiguous. On this view, slurs that are not appropriated have univocal derogatory contents while appropriated slurs are ambiguous (or polysemous) between a derogatory and a non-derogatory content.[4] For example, Richard says that "there is a case to be made that in appro-

---

[1] Mentioning and of course using a slur can cause offense and other serious harms. The practice of mentioning slurs rather than using phrases like 'the N-word' or 'b----' is standard practice in the philosophical and linguistics literature on slurs and pejoratives. Camp argues in support of explicit mentions of slurs stating that "we can understand slurs' actual force only by considering examples where we ourselves experience their viscerally palpable effects" (2013: 331). Bolinger argues that the choice to mention a slur may be "associated with tamer (though not benign) attitudes, ranging from simple insensitivity to perverse pleasure at saying discomfiting words, and disregard for the risk of encouraging derogating uses of the slur" (2017: 452). I take good academic writing to require clarity and I hold that it should be sensitive to various readers' experiences, encourage inclusiveness, and avoid derogation. In an attempt to meet both I will minimize mentions of slur. I will also mention only one slur—'bitch'. I do so to provide a more concrete account with specific linguistic data, while avoiding (what might be perceived as) gratuitous mentions of multiple slurs. I use 'bitch' because it is an expression that is widely held to be in the process of appropriation and its use by individuals outside the target groups is less offensive than some other examples of appropriated slurs. I hope by limiting the number of slurs mentioned, offense and other harms can be minimized and clarity can be maintained.

[2] Although see Anderson and Lepore (2013a, 2013b), Bolinger (2017), and Nunberg (forthcoming) for views on which the derogatory features of slurs are based on non-semantic prohibitions, on term and negative stereotype endorsement, and on conversational implicatures, respectively.

[3] The term 'appropriation' is most commonly used to denote this phenomenon. While I think 'reclamation' is slightly better (partly due to the negative associations 'appropriation' has from discussions of cultural appropriation), I will follow standard usage and use 'appropriation' from here.

[4] There are, of course, other options for handling appropriation. For instance see Anderson (forthcoming), Bianchi (2014), and Lycan (2015). I am not arguing that positing an ambiguity is the best route to handle appropriation. It is, however, a common route and one with an apparent problem that I offer a solution to here.

priation there [is] a change in meaning" (2008: 16). Potts claims that "when lesbian and gay activists use the word 'queer', its meaning (and its expressive content) differs dramatically from when it is used on conservative talk radio" (2007: 10). Jeshion suggests that ""queer" became semantically ambiguous upon appropriation" (2013: 326). Hom states that appropriation "alters [a slur's] meaning for use with the group" (2008: 428). Whiting argues that in appropriation "the expressions bear a different meaning than they would otherwise bear, at least insofar as (once appropriated) they no longer conventionally implicate the relevant negative attitude" (2013: 370). And Saka ties meaning change into the very definition of appropriation. He states that appropriation is when a "victim group attempts to change the meaning of some term" (2008: 42). The strategy of positing ambiguity to account for appropriation is clearly widespread.

While positing ambiguity[5] is a natural move for a proponent of content-based view of slurs, a problem looms. The theorist positing multiple meanings needs to explain why at least in many cases of appropriation one of the meanings the term comes to have can only be expressed by the term when it is used by members of the targeted group. For instance, not just any speaker can use the N-word or 'bitch' to mean something friendly or positive. Other ambiguous and polysemous expressions do not place restrictions on who can use them to express one or the other of their contents. For instance, anyone can use 'duck' to denote an aquatic bird or a crouching action. Anyone can use 'bank' to mean financial institution or side of a river. Anderson and Lepore pose the problem stating "[a]mbiguity fails to explain why non-members cannot utilize a second sense. If it were *just* a matter of distinct meanings, why can't a speaker opt to use a slur non-offensively?" (2013a: 42). The worry they target is what I call the Appropriation Worry.

> *Appropriation Worry:* Content-based views that posit an ambiguity to account for appropriation cannot account for why only members in the target group (and perhaps others with "insider" status) can use an appropriated slur to express a non-offensive/positive meaning.

Later Anderson states that "at the very least, [an ambiguity view] must be paired with an additional explanation, one that details some kind of rule-like structure that governs access to the appropriated sense"

---

[5] From here I use 'ambiguity' to mean 'ambiguity or polysemy'. One way of understanding the difference between ambiguity and polysemy is as follows: ambiguity involves separate words that are orthographically or phonologically identical. Polysemy involves a single word with multiple meanings. The ways "classic" ambiguous expressions—like 'bank'—and examples of polysemous expressions—like 'bottle'—pattern on Zwicky and Sadock's (1975) identity tests give some evidence that ambiguity and polysemy are distinct phenomena. For additional discussion and treatments of polysemy see, for example, Falkum and Vicentea (2015).

(forthcoming, 6). That is the task I take up here. I develop a solution to the Appropriation Worry that could be adopted by any content-based theorist. The Appropriation Worry does not spell defeat for content-based views that posit ambiguity. I argue for a solution to the Appropriation Worry on which appropriated slurs are ambiguous between a derogatory content and an appropriated content that involves a plural first-person indexical. I show how the semantics of plural first-person indexicals can be put to use in the solution and account for why only target group members (and perhaps also those with "insider status") can use appropriated slurs to express their positive contents. I also argue that the solution is motivated by the process of appropriation and that it helps to explain how appropriation achieves its aims of promoting group solidarity and positive group identity.

I begin in *1* by laying out the range of content-based views. Then, in *2*, I consider the process and aims of appropriation. I also canvass data that a solution to the Appropriation Worry must capture and explain. In *3* I consider the ways that indexicals can be sensitive to a speaker's position in physical and social space. In particular, I argue that a speaker's social group memberships can constrain or determine the content of her uses of plural first-person indexicals. In *4*, I sketch a version of an ambiguity account that includes a first-person plural pronoun in the appropriated content. I illustrate the account by considering the case of 'bitch'. I argue that the account is well motivated and can solve the Appropriation Worry. In *5*, I briefly compare the view developed to other treatments of slurs involving indexicality. I summarize the arguments and draw conclusions in *6*.

Before continuing, three clarifications are in order. First, I am not arguing that slurs must be accounted for by a content-based view or that ambiguity should be posited to account for appropriation. That is, my aim here is not to defend content-based views or the ambiguity solution to appropriation. Rather, I am arguing that the Appropriation Worry can be solved. It is not a reason to reject either content-based theories or an ambiguity view of appropriation.

Second, recall that the Appropriation Worry requires a content-based theorist to account for why only members in the target group can use an appropriated slur to express a non-offensive/positive content. The only cases of appropriation the worry targets are those in which an appropriated expression has both its original derogatory meaning *and* a new (likely) positive meaning. The ultimate aim of at least some cases of appropriation is plausibly to completely obliterate the derogatory meaning of a slur and supplant it was a positive or neutral meaning.[6] Some former slurs, e.g., 'tar heel' and 'Whig', are arguably no longer derogatory and can be used by any speaker to express something

---

[6] In some cases of appropriation , the aim might be to reclaim a slur solely for in-group usage. I do not claim that the only aim in appropriation is to supplant derogatory meaning with a new positive or neutral meaning. Thanks to Robin Jeshion (pc) for emphasizing that appropriation projects might have different aims.

non-derogatory. The expression 'queer' plausibly has an appropriated meaning that is expressible by anyone tokening it (including people who are cis-gender and heterosexual). There are now Queer Studies departments and the LGBTQ movement has included 'queer' in its acronym, which can be appropriately used by any speaker.[7] Given that my aim is to provide a way for a content-based theorist to avoid the Appropriation Worry cases in which a new meaning has supplanted an old will not be considered. Here I focus only on expressions that are often treated as having two meanings—one positive meaning expressible by the appropriated slur only by in-group members and one negative meaning expressible by the slur by any users.

Third, there is ongoing disagreement about whether appropriation can lead to positive effects and, so, whether appropriation should be attempted. Some (e.g., Asim 2007 and Kleinman et. al. 2009) argue that so-called appropriated uses express self-hatred and reinforce racism, heteronormativity, and patriarchy. For instance, Asim argues that "[a]s long as we [Blacks] embrace the derogatory language that has long accompanied and abetted our systematic dehumanization, we shackle ourselves to those corrupt white delusions" (2007: 233). In contrast, Kennedy argues that "[s]elf-hatred … is an implausible explanation for why many assertive, politically progressive African Americans" use the N-word, rather they "in their minds at least" use the expression "not in subjection to racial subordination but in defiance of it" (2003: 37). There are complicated social, political, and moral issues surrounding appropriation that cannot be addressed here.[8] If appropriation is impossible, there is no Appropriation Worry to solve. Given the widespread view that appropriation does occur and given the aim to provide a solution to the Appropriation Worry here I adopt the following assumptions: (i) appropriation is possible and (ii) appropriation can produce lexicalized contents that are positive rather than derogatory.

1. *Content-Based Views of Slurs*

I categorize any view that takes slurs to include a derogatory component in their conventional lexical meaning to be a content-based view. The derogatory component could be a conventional part of truth-conditional meaning, an additional non-truth-conditional expressive content, a presupposition, or a conventional implicature. Since each of

[7] Although 'Q' is also sometimes taken to stand for 'questioning'. Thanks to Matthias Jenny for bringing this point to my attention.

[8] For instance, see also Langton, Haslanger, and Anderson (2012) and Wodak and Leslie (forthcoming) for worries that generic generalizations about racial groups essentialize (or can easily be interpreted as essentializing). One component of social justice projects involves showing that there are not innate racial essences that manifest in intellectual, moral, and personality traits. If a slur, even used to express its positive appropriated meaning, is used in a generic generalization their worries about essentializing apply. See Tirrell (1999) for discussion of arguments for abolishing versus reclaiming slurs.

these four views takes slurs themselves to conventionally express (in some way) a derogatory content, I classify all as content-based views. There are, of course, other ways to classify views. For instance one might argue that presuppositional and conventional implicature views are pragmatic, while only views on which slurs have derogatory truth-conditional content should be considered "content views".[9] Given that (a) the views just listed all include something derogatory in the lexicalized content of slurs, and (b) the Appropriation Worry is a worry for any view that take slurs to conventionally express (in some way or other) a derogatory content and posits ambiguity to account for appropriation, this classification schema will be useful for our purposes. I will sometimes use the expressions 'semantic' and 'meaning'. I intend these expressions to be understood in ways that are neutral between the various content-based views to be discussed. In particular, in their uses here I do not take 'semantic' or 'meaning' to require more than conventional lexical content. Next, I briefly lay out versions of the four content-based views.

First, are what I call truth-conditional content views (e.g., Hom 2008, 2010 and Hom and May 2013). On these views a slur for a group of people that can also be referred to by a neutral counterpart N means something like *N* and worthy of contempt for being so*. Hom argues that multiple stereotypes are part of the truth conditions of utterances that include slurs. On his view, a slur S with neutral counterpart N has a complex truth-conditional semantic value of the form 'ought to be subject to $p^*_1 + \ldots + p^*_n$ because of being $d^*_1 \ldots d^*_n$ all because of being N*' (Hom 2008: 431). The properties $p^*_1 + \ldots + p^*_n$ are deontic prescriptions about how the person should be treated given the negative properties derived from racist (or sexist, or homophobic, or …) practices given by $d^*_1 + \ldots d^*_n$ and N* is the semantic value of *N*. For instance, he states that "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: 431). On truth-conditional content views, the meaning of a slur and its neutral counterpart are not truth-conditionally equivalent as the latter includes a (perhaps complex) derogatory component.

The remaining three classes of views take slurs to be truth-conditionally equivalent to their neutral counterparts, but take the expres-

[9] For instance Sennett and Copp (2015) classify *only* views like what I call the truth-conditional content views to be content views. Bianchi (2014) takes strategies that rely on conventional implicatures or presuppositions to be pragmatic, seemingly going against classifying the four strategies I take to be content-views together. However, she notes that the "(semantic or pragmatic) status" of conventional implicature and presuppositional views "is far from settled" noting that Potts (2007) and Camp (2013) classify their views as semantic. Hom (2010) argues for a six way classification.

sions to have another dimension of meaning that captures its offensive and derogatory content. According to presuppositional content views (e.g., Schlenker 2007 and Cepollaro 2015), slurs contribute only their neutral counterparts to truth conditions and they presuppose something derogatory. The derogatory component is lexicalized or "built into" the meaning of the slur so that all utterances of slurs carry a negative presupposition.[10] According to presuppositional accounts slurs are akin to expressions like *quit*. Utterances of "Andy quit smoking" and "Andy did not quit smoking" both presuppose that at a time preceding the time of utterance Andy smoked. Similarly, presuppositional content views hold that "Anne is a bitch" and "Anne is not a bitch" both presuppose something like *Anne is despicable for being a woman* or *the speaker believes people who are women are worthy of derogation*.

Conventional implicature content views (e.g., Potts 2005, Williamson 2009, Whiting 2013, Lycan 2015) hold that slurs contribute the equivalent of their neutral counterpart to the truth conditions (e.g., the equivalent semantic content that 'Black', 'gay', 'Jewish', or 'woman' express), and conventionally implicate something negative. The notion of conventional implicature goes back to Grice (1975). He argued that expressions like 'but' carry implicatures that are not calculated based on a specific conversational context and conversational maxims. Instead, they are lexically conventionalized. For instance, "Nwando is poor but honest" truth-conditionally expresses the same content as "Nwando is poor and honest" but conventionally implicates that it is unusual to be both poor *and* honest. The implicature is not dependent on specific contextual or conversational features, but is part of the import of the word 'but' itself (albeit not a part of its truth-conditional content).[11] Applied to slurs conventional implicature content views hold that slurs conventionally implicate something derogatory. For example, Williamson (2009) argues that "A is a Boche" means that A is a German and conventionally implicates that A is cruel.

Finally, expressivist content views (e.g., Saka 2007, Richard 2008, Jeshion 2013) hold that the truth-conditional contribution of a slur is identical to its neutral counterpart, but an additional expressive element or content is also conveyed. For example, Jeshion argues that in addition to their truth-conditional content, slurs have an expressive component that expresses "contempt for members of a socially relevant

[10] There are multiple views of the way presuppositions work. Most take presuppositions to be at least partly semantic in nature. There are, however, purely pragmatic accounts of presupposition (see, e.g., Stalnaker 1974). On such views presuppositions are understood wholly in terms of conversational contexts. The views that take the derogatory content of slurs to be presuppositional take the presuppositions to be carried by the lexical item (i.e., the slur itself) so would be classified as (at least partially) semantic accounts of slurs. They can, therefore, legitimately be categorized as content-based views.

[11] While 'but' is a classic example used in discussions of conventional implicature, not all agree that it involves conventional implicature (see, e.g., Potts 2005).

group on account of their being in that group or having a group-defin-ing property" (Jeshion 2013: 316).[12]

The preceding discussion evidences that many theorists hold con-tent-based views of slurs. Each of these theories needs to be paired with an account of appropriation. As we saw above many theorists pos-it ambiguity to account for appropriation. In such cases, they also need a response to the Appropriation Worry. Next, I turn to data about the process and results of appropriation.

2. *Appropriation Data*

In ordinary cases, one speaker using an extant word in a non-standard way does not alter its meaning. It is only through multiple uses of the word in what was a non-standard way that a word's meaning can change or that a word can come to have an additional meaning. Appropriation is a process that takes time, multiple uses, and multiple speakers.[13] Cases of appropriation involve a slur being taken up by members of the group the slur targets for positive in-group usage—thereby undermin-ing the derogatory content of the slur.[14] In this section I consider what is required for appropriation. Then, I turn to several data points that must be captured, and ideally explained, by any adequate treatment of appropriation.

Appropriation is social and political. It works to emphasize and con-struct group identity and to promote group solidarity. Tilly argues that "social movements stand out for their emphasis on identity assertion" and highlight that individuals with the identity are "worthy, unified, numerous, and committed" (Tilly 2002: 121). The famous slogan from

---

[12] Jeshion includes a third component in her semantics of slurs as well—what she calls an identifying component. She states that "as a matter of their semantics" slurs "are used so as to signal that being [for example] Jewish, Chinese, [B]lack, gay, a prostitute identify what its targets *are*" (Jeshion 2013: 318, *emphasis original*).

[13] There might be exceptions to this. For instance, someone with a lot of authority might be able to appropriate a term with a single widely heard utterance. For instance, Robin Jeshion (pc) suggested that Obama using 'Obamacare' in a positive way might have been enough for it to be appropriated. In a tweet on March 23, 2012 Obama said "Happy birthday to Obamacare: two years in, the Affordable Care Act is making millions of Americans' lives better every day." While Obama's tweet likely played a large part in the appropriation effort—a campaign involving emails from top White House officials, a website, and a hashtag were also created to help reclaim the term. So, it is not clear that even when an utterer has authority and a large audience that a single use is enough to reclaim a derogatory expression.

[14] At least usually it is in-group usage that results in appropriation. Beaton and Washington (2015) discuss a case of 'favelado' a slur in Brazilian Portuguese referring to individuals who live in slums, that has been appropriated in the context only of fans of the soccer team Flamengo. The term was used by opposing teams to derogate Flamengo fans and is now used within the group of Flamengo fans as a term of solidarity. Since Flamengo fans were taken by opposing team fans to be referents of 'favelado' I take the example to be very similar to appropriation by in-group members.

the activist group Queer Nation—"We're here! We're Queer! Get used to it!"—is a prime example. In the chant a slur is appropriated as a means of self-identification and it is demanded that being queer be normalized and respected. Croom states that a sense of solidarity can be fostered by in-group uses of a slur and that this "can help speakers signal to each other that they are not alone and that others like them share in their pains, perspectives, and history of prejudices" (Croom 2011: 350). Hom argues that target group use of a slur can serve as "a means of in-group demarcation to bring members of the targeted group closer together" (Hom 2008: 428). By self-labeling as Ns members of a target group emphasize their identity as oppressed people and reinforce that there is a shared position from which political demands can be made.

In appropriating a slur for self-identification, members in the target group might do more than emphasize their shared history; they can also *construct* identities involving in-group norms about ways to act, dress, communicate, and so on.[15] Appropriation is part of a project that emphasizes that these are strengths of *ours*; this is how *we* dress, talk, and act; *we* have persevered. Rahman states that since the N-word is "a self-selected term for naming members of the group, it is an apt subject for yielding insight into the way that at least some African Americans see themselves and their community" (Rahman 2012: 139). She goes on to say that, at least in some cases, its use contributes to a person's "presentation of self within the community" (2012: 140). Appropriation of a slur can emphasize and construct identity and promote solidarity. These are key features of identity politics and social justice movements.

In addition to identity-building and solidarity-promoting, appropriation works to remedy power imbalances and remove weapons from oppressors. Hornsby states that old derogatory meanings are not "brushed away: they are subverted" (Hornsby 2001: 134). This is political action. In appropriation oppressed people rise up to confiscate a linguistic tool that functions to reinforce oppressive social structures. Appropriation involves pointing out oppressive social and cultural norms and working to counteract them. McConnell-Ginet argues that "to use *queer* both to affirm difference from heterosexual norms and to refuse efforts to eliminate or reduce such differences is to claim a kind of 'mastery', to refuse the conjunction of abuse and attribution of homosexuality so prominent in the … history of the word *queer*" (McConnell-Ginet 2011: 254). The slur being appropriated is altered to reject the combination of negativity with social group identity. Kennedy argues that in reclaiming slurs marginalized groups "have thrown the slur right back in their oppressors' faces" (Kennedy 2003: 38). Finally, Hom (2008) argues that appropriation involves taking back a powerful tool of discrimination in an effort to remove the offensive power of a slur. Each of these theorists stresses that appropriation is an action that

---

[15] I thank Rachel McKinney for stressing the point the appropriation does not just express, but constructs identity.

aims at reconfiguring power imbalances by laying claim on a tool of the oppressors.

A recent study provides evidence that slur appropriation can have measurable effects at least on perceived power imbalances. Galinsky, et al (2013) found that self-labeling with a slur increases an individual's sense of power and increases an observer's evaluation of both the self-labeler's power and the power of the target group. Moreover, they found that self-labeling led to decreased perceptions of negativity in the slur that was used to self-label. To summarize, appropriation is a process that is social and political in at least the following four ways: (i) it emphasizes and constructs group identity (ii) it promotes solidarity (iii) it works to remedy power imbalances, and (iv) it takes tools from oppressors.

Bianchi (2014) argues that appropriation is not always social or political. She argues that while appropriation can occur in social and political contexts it can also occur in what she calls "friendship contexts" when one is joking amongst friends. Even in the context of a joke, I argue that appropriation is social and political. Slurs are part of a social-historical context of oppression and power imbalance. Even if appropriation begins as a part of a joke, it always relies on socio-political features of slurring expressions. So, it is always social and political.

To see why, consider pejoratives like 'asshole', 'dick', or 'jackass'. These expressions are frequently used in the context of jokes. If joking around using a term was sufficient for appropriation, these expressions would plausibly have appropriated positive meanings. Yet, it is clear that they do not. For instance, even though 'asshole' is often used ironically or in a joking manner, it has not gained a new positive meaning. If appropriation is always tied to the social and political as a part of a process to undermine oppressive structures and stereotypes, we can make sense of this data. Expressions like 'asshole' do not target groups of people that are socially oppressed. They are pejorative, but not in the same way that slurs are. Words like 'jerk' are not, as Tirrell states, "tied to a rich structure of other social practices" in the way that slurs are (Tirrell 1999: 62). Even when target group members use a slur as a joke, their use is social and political given the way slurs are tied to social practices, stereotypes, and power imbalances. Joking is not enough to appropriate pejoratives that are not tied to socio-political structures. This evidences that appropriation is social and political. Slurs' social and political nature makes them ripe for appropriation while other non-group denoting general pejoratives are not.[16] Next, I turn to examining linguistic data about appropriated slur use.

Individuals outside a racial, ethnic, gender, sexual-orientation, or other group targeted by a slur cannot use an appropriated slur to ex-

---

[16] This is not to say, of course, that other pejoratives are not subject to changes in their meanings. Meaning changes can occur in any expressions, but not all changes of meaning are cases of appropriation. Thanks to David Plunkett for useful discussion that led me to consider pejoratives that do not target groups.

press its appropriated meaning. Or, at least, those outside the target group can use a slur to express its appropriated meaning only in very rare circumstances. For instance, someone might gain "insider" status in a target group that allows her or him to use the slur to express its appropriated content. Kennedy argues that there are relationships between a black person and a white person that are such that "the white person should properly feel authorized, at least within the confines of that relationship, to use the N-word" (Kennedy 2003: 42). This gives us our first datum—and one half of the Appropriation Worry.

> *Outsider Usage*: For an appropriated slur *S* which targets group *g*, individuals outside of *g* cannot (or perhaps very rarely with "insider" status) use *S* to express its appropriated content.

Our second datum is that individuals in the target group are able to use appropriated slurs to express their appropriated content. This and *Outsider Usage* are the data that the Appropriation Worry trades on. It should also be noted that it is possible for a member of the targeted group to use an appropriated slur to express its original derogatory meaning. For instance, a woman can use 'bitch' to express an unappropriated negative content. Being in the target group does not force someone to use a slur to express its appropriated positive content.[17] We can formulate this data as:

> *Insider Usage*: Members of group *g* targeted by an appropriated slur *S* can use *S* to express its appropriated content or its original derogatory content.

Next, I turn to developing a semantics of appropriated slurs that captures *Outsider* and *Insider Usage,* solves the Appropriation Worry, and which takes seriously the features and aims of appropriation discussed above.

3. *Indexicals and Social Identities*

We are physically located in space and time. We are also socially located in a "space" of social groups. We are members of unions, departments, boards, and clubs. We are in social kinds—like the kinds New Yorkers, Canadians, philosophers, immigrants, women, and Latinos. In order to solve the Appropriation Worry, I argue that one's position in social space—that is what groups one is a member of—can affect what one can express with certain expressions.

Indexicals are expressions whose contents vary from context to context depending on, for instance, a speaker's location in space and time. In this section I argue that first-person plural indexicals are sensitive to social position. That is to say, they are sensitive to what social groups

---

[17] Anderson (forthcoming) also argues that appropriated slurs can be used to by target group members to express something negative, although he argues that with negative in-group uses of the N-word the attitude expressed is different from that expressed by bigots' uses of the N-word.

a speaker is in. I argue that first-person plural indexicals' sensitivity to position in social space places restrictions on what a speaker can use 'we' and 'us' to denote.

The section proceeds as follows. I begin by considering familiar cases in which a speaker's physical location determines what is expressed by (or restricts what she can express by) indexical expressions like 'here' and 'now'. Then I consider cases of pronouns with gender features. Finally I consider cases of singular and plural first-person pronouns. I will rely on the view that a speaker's position in physical and social space can restrict or determine what she can mean by an indexical in my solution to the Appropriation Worry.

As is well known, a speaker's position in space and time can determine or constrain the content of her uses of indexical expressions like 'here' and 'now'. According to Kaplan (1989) indexicals are directly referential expressions with both a character and a content. Characters are meaning rules that can be modeled by functions from contexts to contents. Contexts are parts of the world that can be modeled as ordered tuples including at least the speaker, addressee(s), time, and location. Contents are objects (e.g., places or people). Let's work through an example. The character of 'here' can be modeled by a function from contexts (modeled as ordered tuples) to the location in the ordered tuple. A token of 'here' uttered in a context in which New York City is the contextually specified location has NYC—that object—as its content. The character is used to deliver a content, but is not part of what is expressed. 'Here' directly refers to a place on each occasion of use. Similarly, according to Kaplan an utterance of 'now' directly refers to the time of the utterance (the time in the context), not, say a time 10 days later or 3 years earlier.[18] Even if one rejects a direct reference theory of indexicals, any account of expressions like 'here' and 'now' must account for the way the utterance context determines what is meant (or constrains what a speaker can intend) on an occasion of use.

These familiar examples show that what an expression can be used to express can depend on a speaker's location in space-time. They do not, however, appear to be sensitive to any specific features of the speaker. For instance, anyone in New York City can use 'here' to refer to it. The sort of sensitivity appropriated slurs manifest is more specific. Whether a slur can be used to express its appropriated content does not vary with movement through physical space, rather it is the speaker's social group memberships—whether he is a man, Black, heterosexual—that matters. We need evidence to show that expressions can be sensitive to social group membership in order to have the re-

---

[18] There are issues with the boundaries of locations in the case of 'here' and 'now'. For instance, is the contextual location the room in which an utterance is made? The city? State? Country? There are also issues with recordings, like an answering machine, that might require a more complicated treatment. See Kaplan (1989) for an initial presentation of the case. See Cohen and Michaelson (2013) for a range of possible responses to cases like these.

sources to solve the Appropriation Worry. I argue that gender features provide some evidence for social position sensitivity.

Gender features on expressions often require that their denotations have a particular gender. For instance, the gender features on the English pronouns 'she' and 'he' almost always restrict their denotations to feminine and masculine entities, respectively.[19] Gender features can affect interpretation and felicity. Consider the following examples:

1. Laura met Chris at noon. She likes him.
2. Laura met Chris at noon. He likes her.
3. Laura walked into the house. ??He saw that the lights had been left on.

The gender features on the pronouns in 1 and 2 deliver different interpretations about who likes whom. In a null context 3 might be interpreted as infelicitous given the masculine feature of 'he' and that 'Laura' is stereotypically a name for a woman. Alternatively, it might provide the information that Laura identifies as a man.[20] Gender features on pronouns can be sensitive to a denotation's gender and can affect interpretation.

It has been argued since at least de Beauvoir (1949/2011) that gender is social. De Beauvoir famously argued that "[o]ne is not born, but rather becomes, woman" (1949/2011: 330). Numerous accounts of gender as a social feature have been proposed. All share at their core the claim that gender is not wholly natural. Gender does not, for example, depend solely on physiological features like chromosomes or reproductive organs. Haslanger argues that "[g]ender categories are defined in terms of how one is socially positioned, where this is a function of, e.g., how one is viewed, how one is treated, and how one's life is structured socially, legally, and economically" (2000: 38). Butler (1990) argues for a performative conception of gender on which gender is performed through repeated actions (e.g., wearing lipstick and crossing one's legs when seated). For our purposes, a specific account of the social nature of gender is not required. Given that it is overwhelmingly held that gender is social, I will adopt that view here. We will see below several examples that show ways gender, and so, something social, influences what one can express by some particular expression.

So far I have argued that (i) gender features on pronouns restrict interpretation and affect felicity and (ii) that gender is social. Given (i) and (ii) it follows that some expressions are sensitive to social features. This data does not yet show that what a speaker can express with an

---

[19] Of course, grammatical gender systems do not always require a denotation to be gendered in a particular way. That is, grammatical gender is not always aligned with semantic or what is sometimes called "natural" gender. Ships are not literally women, although they are often referred to with tokens of 'she'. Gardens do not have a gender, but in Spanish 'el jardín' is masculine. However, some grammatical gender clearly affects interpretation and felicity.

[20] Thanks to Nicole Dular for offering this alternative judgment.

expression is constrained by her social position. For instance, men, women, and people who are gender-non-conforming can all use 'she' to denote a woman. Gender features of third-person pronouns do not restrict *who can use* an expression to express some particular content, but rather restrict *whom the expression can denote*. We have evidence that a speaker's physical location in space-time can affect what she can express with a certain linguistic item. We have evidence that a social feature, namely gender, can affect interpretation and felicity. We do not yet, however, have evidence that what content a speaker *can express* with a certain expression is sensitive to her social features. Next I offer two arguments to show that social features can affect what a speaker can express with certain expressions. The first involves first-person gender marked Japanese pronouns. The second focuses on uses of first-person English indexicals.

Japanese includes first-person pronouns that are gender specific and that vary in degrees of femininity/masculinity and formality. The felicity of a pronoun depends on certain features of the speaker and the context (e.g., the relationship between the speaker and addressee). Even though pronouns have specific gender features, there is variation in the pronouns speakers use to refer to themselves. In a longitudinal study of middle-school children, Miyazaki (2004) found that popularity affected which pronoun a child used. For instance, she found that popular boys used the masculine pronoun 'ore' while unpopular boys used 'boku'. One unpopular boy said that he "wouldn't sound cool at all if [he] used *ore*" (reported by Miyazaki 2004: 256). She also reports that unpopular boys who try to use 'ore' are often bullied or beat up. Girls who use 'ore' are reportedly taken to be rebellious or "crazy", while boys who use feminine first-person pronouns are reported to be homosexual or transgender. Miyazaki's data supports the view that a speaker's gender, identity and social status affect pronoun choice and interpretation. We now have preliminary evidence that what a speaker's social position affects what she can express by some expressions.

English data, particularly involving first-person plural pronouns, provides additional evidence that what one can express with a pronoun depends on one's social position. 'We' can have uses that are sensitive to a speaker's role or group memberships. For instance, consider the following:

4. [Said by a woman] We are less likely to contract the disease than men are.

4'. Women are less likely to contract the disease than men are.

5. [Said by a child of civil rights activists] If my parents hadn't been born, we would be even worse off than we are now.

5'. If my parents hadn't been born, Black people would be even worse off than Black people are now.

Nunberg (1993) offers example 4 and argues that it is equivalent to 4'. If a man were to utter 4 it could not be synonymous with 4'. What

a speaker can express with 'we' is constrained by the group(s) the speaker is actually in. Suppose 5 is uttered by a child of Martin Luther King, Jr. and Coretta Scott King. That speaker might use 5 to express 5'. In contrast, someone white could not use 5 to express 5'. Again, a speaker's group memberships can restrict what she can express by an utterance of 'we'.

Examples 4-5 include adverbs of quantification ('likely') and modals, but 'we' can denote a group that is not individuated merely in terms of extension in other contexts as well. For instance, consider the following:

6. [Said by a Muslim Congressperson] We are rare.
7. [Said by a woman] We get paid less than men for doing the same work.
8. [Said by a Transwoman] We are being murdered without prosecution or protection by hate crime laws.
9. [Said by someone Black] So many forces in American life are telling us that our lives don't matter, that our lives are expendable, that when we are killed when we're unarmed that we can't get justice for that.[21]

The tokens of 'we' in 6–9 are naturally interpreted as synonymous with 'Muslim Congresspeople', 'women', 'transwomen', and 'Blacks', respectively. In these cases, 'we' is used to refer in ways that are strikingly similar to bare plural expressions in generic generalizations. The predicate in 6 ('rare') has been classified as a direct-kind predicate. Direct-kind predicates are overwhelmingly held to take kinds, rather than individuals members of a kind, as arguments.[22] For instance, while no person is rare, Muslim congresspeople and other social or natural kinds can be rare. Sentences 7-9 are examples of characterizing generics. Characterizing generics specify some characteristic that is common or striking although perhaps not universal in the group. For instance, 7 could be true even if some women are not paid less than any men for doing the same work. The speakers of 8 and 9 have not been killed. Nevertheless, 8 and 9 could be (and, unfortunately, are) true.[23]

The first-person pronouns in 6-9 are not being used to denote some purely extensionally defined set of individuals. Rather, they are used to denote social groups that share some feature(s). These examples show that an overt adverb of quantification or modal is not required to deliver an interpretation of a plural first-person indexical that is sensitive to a property.

Examples 4-9 show that 'we' can be used to refer to a group that is not just a set, sum, or plurality of members, but a group that is speci-

---

[21] Quote from John Legend, April 26, 2015 CNN *State of the Union*.

[22] See, e.g., Krifka, et. al. (1995) and Leslie and Lerner (2016).

[23] There is widespread debate about the correct semantics for generic generalizations. See, e.g., Carlson (1977), Krifka, et. al. (1995), and Leslie and Lerner (2016).

fied in terms of some property (or properties). 'We' does not straightfor-
wardly fit into a Kaplanian direct reference theory of indexicals. One
can be a member of many groups. One is in a member of groups based
on race, gender, nationality, sexual orientation, and religion. One is
a member of a family, cliques, teams, clubs, and departments. In a
conversation one is a member of a group of speakers and interlocutors.
There is not just one group that any speaker's use of 'we' can denote. In
modeling contexts in the way Kaplan does, there is not one element in
the tuple that is *the* group that contains the speaker. Moreover, there
cannot even be one group that is the contextually salient group that
contains the speaker. For example, in a single context the speaker
might use 'we' to pick out her immediate family, the group containing
herself and her interlocutor, and the group of all Americans. 'We' can-
not be a pure indexical in the sense Kaplan argued 'I' was. Speaker in-
tentions, broader conversational goals, and other features are required
to fix the denotation of 'we'.

Moreover, uses of 'we' are often not rigid in the way Kaplan argues
uses of indexicals are. For instance, recall 6, repeated below:

6. [Said by a child of civil rights activists] If my parents hadn't been
born, we would be even worse off than we are now.

If one holds that a person's origins are essential, the speaker of 6 will
not exist in the worlds at which 6 is evaluated. Further, 6 might be true
and denote a group at those worlds. 'We' is not rigid, or is not straight-
forwardly rigid in the way 'I' has been argued to be.[24]

Consider one more case to emphasize the point.

10. We might have been liberals.

Nunberg (1993) imagines 10 uttered by a Supreme Court justice in
the context of a discussion of what the make up of the Supreme Court
might have been given different results in some salient presidential
elections. He argues that in that context 10 need not mean that the ac-
tual Supreme Court justices might have been liberals, but rather that
the justices who would have been appointed would have been liberals.
Again, the denotation of 'we' is allowed to vary across worlds in a way
that, at least prima facie, is at odds with rigidity. The data canvassed
provides evidence that what a speaker can express by 'we' is restricted
by her actual roles and memberships in various groups.

Tokens of 'we' denote groups. Some groups are specified in terms of
properties, like *being women*, rather than merely in terms of having

---

[24] One could argue that 'we' rigidly picks out a group that can vary in members
across times and worlds. Even if this is the case, it is rigid in a way that is much
less straightforward than 'I'. Moreover, to avoid making rigidity trivial, the account
of rigidity that allows for group variation must not entail that all terms are rigid.
For instance, it should not entail that, e.g., 'the tallest people' rigidly picks out a
group that varies across times and worlds. See Devitt (2005) and Besson (2010) for
discussion of rigidity and natural kind terms that could inform the viability and
structure of an account of rigidity for 'we'.

particular individuals as members. The data in 4–10 shows 'we' can denote a group that is not just specified in terms of extension. The semantics of 'we' can be sensitive to properties.[25]

4. *Solving the Appropriation Worry with Plural Indexicals*

Recall that according to many content-based views, upon appropriation a slur becomes ambiguous. The Appropriation Worry argued that positing ambiguity failed to account for who could use a slur to express its positive appropriated content. We now have the resources to give a solution to the Appropriation Worry. In sketching the solution, I will use 'bitch' as a case study. 'Bitch' and 'bitches' are plausibly forms of a slur that has an appropriated meaning and that retains its original derogatory meaning. My aim here is to argue for a solution to the Appropriation Worry that provides the resources to supplement any version of a content-based view of slurs. While the specifics of content-based views vary, all take slurs to express something derogatory through either truth-conditional content, presupposed content, a conventional implicature, or an additional expressive content. Here, I will present the solution within the framework of a conventional implicature content-based view. I do so not because I intend to defend a conventional implicature view (or any other content-based view). Rather, I do so in order to represent the solution to the Appropriation Worry more explicitly while avoiding gratuitous repetition. The view I offer is available to any proponent of a content-based view with suitable adjustments.

While my aim is not to argue for any precise definitions of slurs or appropriated contents, nevertheless, it will be useful to consider fairly explicit examples for the slur under consideration. The specifics could be adapted to, for instance, include specific stereotypes or other features.[26] Consider a sentence containing 'bitch' that is ambiguous be-

[25] Nunberg (1993) argues against Kaplan's direct reference theory of indexicals using data like 5 and the following: [Said by a condemned prisoner] I am traditionally allowed to order whatever I like for my last meal.

He takes the sentence to express something like "an inmate on death row is traditionally allowed to order whatever he likes for his last mean". Nunberg calls these "descriptive uses" of indexicals. Moreover, he argues that they must be captured semantically. So, the direct reference theory of indexicals is false. There is widespread dispute about whether Nunberg's view that there are descriptive elements in the semantics of indexicals is correct. See, e.g., Nunberg (1993) and Elbourne (2008) for arguments in favor of semantic accounts of "descriptive uses". See, e.g., Recanati (1993), King (2006), and Hunter (2010) for arguments against semantic accounts of "descriptive uses". Whichever way one sides in the debate about other indexicals my arguments against 'we' being a pure indexical and against a simple story about the rigidity of 'we' provide evidence that its semantics is not accommodated by a Kaplanian direct reference theory.

[26] On the view I sketch here, I take the original content of 'bitch' to be non-indexical and its appropriated content be a plural first-person indexical. I should note, however, that one could take both meanings to be indexical. Perhaps the original content emphasizes that women are other by including 'they', while the appropriated

tween expressing a slur's original derogatory content and its appropriated positive content:

11. I'm going out with my bitches tonight.

On the view under consideration, 11 has two possible interpretations as in 11' and 11" (where 'TC' stands for truth-conditional content and 'CI' stands for conventionally implicated content):

11'. TC: I'm going out with my women tonight.
    CI: They are despicable or lesser than for being women.

11". TC: I'm going out with my women tonight.
    CI: We women are laudable for being women.[27]

The original content is a simple truth-conditional content on which 'bitch' truly applies to someone if the person is a woman. It conventionally implicates that the subject is despicable or lesser. In the appropriated content, the neutral element, 'woman', is retained and the conventional implicature is subverted, implicating that we are worthy for being women. 11' could be expressed by anyone uttering 11. In contrast, the Appropriation Worry traded on the fact that 11" can only be expressed by utterances of 11 by certain speakers. That is, not just anyone who utters 11 can express 11". To express the appropriated meaning, a speaker must be part of the targeted group—in this case the group of women. The indexical in 11" captures why only some speakers can express the appropriated meaning of 'bitches' by uttering 11. To see how the account will go, we need to examine the semantics of 'we'. Before doing so, I want to justify the use of a plural first-person pronoun.

The use of the plural first-person is motivated by the process and aims of appropriation. In Section *2* I argued that appropriation is a social and political action that reinforces solidarity and constructs and emphasizes group identity. 'We' emphasizes that there is a group that is a subject rather than merely a group shaped and created by oppressors. As de Beauvoir argued in using 'we' proletarians and Black people emphasize that they are subjects and "transform the bourgeois or whites into 'others'" (1949/2011: 28). The use of 'we' in these cases emphasizes solidarity. Further, since 'we' is being used to denote a group with members beyond the immediate conversational context the professed solidarity goes beyond the participants in the utterance context.

Relying on a plural, rather than singular, first-person pronoun is also motivated by the following observation. It is possible for a speaker

meaning emphasizes us and includes 'we'. Again, my aim is not to argue for some very specific entries, but to argue that appropriation involves indexicality. Thanks to John Kulvicki for pressing me to think more about this issue. Miščević also considers whether there is an us/them element involved with pejoratives (2016: 138-139).

[27] These are possible contents for a proponent of a conventional implicature view. If one prefers one of the other content-based views, one could take the CIs to be part of the truth-conditions, presuppositions, or expressive content of 'bitch'. Minimal adaptations of the account allow the strategy I advocate to be adopted any of the content-based views.

who uses a appropriated meaning not to think of herself as particularly laudable. That is, it possible for someone who is a woman, Black, homosexual, or in some other group targeted by a slur to use an appropriated slur to express something positive about the group while also questioning their own strengths or worth. For instance, a woman who currently evaluates her creative outputs, job performance, body, or other features negatively, might still use 'bitches' to express something positive about women. For instance, suppose that instead of the appropriated content of 11 being 11" it was 11''':

11'''.   TC: I'm going out with my women tonight.
         CI: I am laudable for being a woman.

If the conventional implicature (or truth conditions, or presupposition, or expressive content) required that the speaker believe of herself that she is laudable, it would be infelicitous or contradictory for a speaker to express a negative self image while using 'bitch' to express its appropriated content. Yet, in appropriation an individual member of a targeted group need not think of *herself* as especially laudable, even while expressing that a group of which she is a member is deserving and admirable. The account I offer, which relies on 'we' in the appropriated content of a slur can allow for this combination of attitudes. To see why consider the following case.

In a certain context, 12 could express a true generalization about women, but one which does not apply to the speaker.

12. [Said by a woman] We get paid less than men, although I am the highest paid person at my firm.

12 is felicitous and easy to interpret. The truth of generic generalizations, like in 12, does not require that every member of a kind satisfy the predicate. This provides evidence that in certain cases the speaker herself might be excepted from the extension of the predicate that takes the denotation of 'we' as argument. Although a full story is needed, it is in principle possible for a speaker to think that she is not laudable, while saying that we women are laudable. Next I argue that *Outsider Usage* and *Insider Usage* can be captured by the proposal being sketched. To fill out the account, we need to look more carefully at the semantics of 'we'.

Pronouns, including 'we', are often taken to carry presuppositions that place requirements on their satisfiers. Heim and Kratzer (1998) argue that number, gender, and person features are presuppositional. For instance, for a token of 'she' to receive an interpretation, there must be a possible denotation that meets the constraint of being a woman/girl (i.e., satisfying that property).[28] Otherwise, it is standardly argued

---

[28] Recall above the claim that gender is social rather than biological. If gender requires being female the distinction between sex and gender is dissolved. In their analysis, Heim and Kratzer require that the property *being female* be met by any possible denotations of 'she' or other feminine pronouns. To account for gender identities that do not match biological sex, here I take the property that must be met

that the token does not receive an interpretation and the entire sentence is neither true nor false.

A use of 'we' denotes a group that the speaker is actually a member of. 'We' is not the mere plural of 'I'. It does not denote a group of speakers speaking in unison. Rather, as Wechsler puts it 'we' denotes "the speaker plus associates" (2010: 377). The solution to the Appropriation Worry I offer here requires that the following additional presuppositions are adopted. First, it requires that tokens of 'we' presuppose that the speaker/writer is a member of a salient (or intended) group *g*. Second, following Heim (1982) on descriptive content presuppositions, when 'we' occurs in phrases of the form 'we Fs', it carries a presupposition that the members of the group, *g*, be Fs.[29] If a presupposition is not satisfied the utterance is infelicitous. If the presupposition(s) are met, 'we' denotes or refers to the group *g*. These provide the necessary resources for a solution to the Appropriation Worry.

In the last section I argued that 'we' always denotes a group, that some groups are specified in terms of properties, and so 'we' sometimes denotes a group that is intensionally specified. I argued that the evidence in 4-10 supports this conclusion. Moreover, the evidence was used to argue that what one can express by an expression can be sensitive to the features of a speaker. I should note, however, that the solution to the Appropriation Worry I offer here could be adopted with weaker commitments. The conventional implicature I offered in 11" includes the noun phrase 'we women' rather than simply 'we'. By using 'we women' a descriptive content presupposition can be appealed to, thereby avoiding taking a stance on whether the semantics of 'we' is intensional.[30]

In 11", 'we women' triggers both presuppositions. The group denoted by 'we' must meet the condition of being composed of women and the speaker must be a member of the group. So, the speaker must herself be a women for 11" to be felicitous. If a man were to utter 11 attempting to express 11" it would be infelicitous. Given the presupposition failure in his attempting to express 11" one might hold that either that he actually expressed 11' or, at least, that he would plausibly be interpreted as expressing 11'. The semantics of 'we women' explains why only members of the target group can felicitously use slurs to express their appropriated contents and helps to explain why utterances of slurs by outsiders, even those who are attempting to use slur to express something positive, are taken to be derogatory or defective.

by a possible denotation of 'she' to be *is a woman or girl*. Whether my modification will allow for non-human animals to be eligible denotations of gendered pronouns will depend on whether non-human animals are boys or girls (they are presumably not men or women).

[29] There might be an additional presupposition that *g* is the the largest group that meets the condition. I am not aiming to give anything like a full theory of 'we' here, so leaving out additional presuppositions is not a fault of the discussion.

[30] Thanks to Matthias Jenny for useful discussion about this point.

The force of the Appropriation Worry has been diffused by the account I have offered. Recall the usage data discussed in *Section 2*.

Outsider Usage: For an appropriated slur S which targets group g, individuals outside of g cannot (or perhaps very rarely with "insider" status) use S to express its appropriated content.

Insider Usage: Members of group g targeted by an appropriated slur S can use S to express its appropriated content or its original derogatory content.

Outsider Usage, sans the parenthetical, has been captured by the account offered. The first disjunct of Insider Usage has also been captured. Let's consider strategies to capture the remaining portions of the usage data.

First consider the possibility that a member of a targeted group can use a slur with its original derogatory content. Here we do not need an independent explanation of how ambiguous expressions are disambiguated. Whatever strategy for disambiguation more generally turns out to be best can be adopted here. For instance, the strategy that best accounts for why an utterance of "I went to the bank" expresses one rather than the other meaning of 'bank' can be used to explain why a target group member's token of a slur has its original or appropriated meaning. Presumably the story will include something about speaker intentions, features of an utterance (e.g., the speaker's tone), features of the conversational context (e.g., relationship between participants, topic of conversation), or other factors. The account I have developed does not force any speaker to use a slur to express its appropriated content.

Next, consider the apparent possibility that in some rare cases a speaker who is not a member of the target group, but who has "insider" status of some sort can use a slur to express its appropriated content. The account I have offered could be supplemented semantically or pragmatically to handle such cases. First, one might hold that insider status in a context allows a speaker to count as satisfying the presupposition that he is a member of the group targeted by the slur. So, at least for the purposes of certain exchanges, the speaker counts as a member of the targeted group. On this view, someone can be a woman (or member of another targeted group) in one context even if they are not in some other context. To develop this strategy a contextually dependent variable could be added to the predicate 'woman' to deliver different extensions in different contexts.

Alternatively, one could argue for a pragmatic account of the apparent ability of insiders to use slurs to express their positive meanings. On such an account the speaker is not *really* able to express the appropriated meaning with his utterance of a slur, but his utterance is not taken to be offensive given his insider status. On this alternative while the speaker literally slurs the target group—given that it was

appropriated for usage only by members of targeted groups—it is clear to all addressees that no offense was intended and that the speaker meant to convey solidarity.[31] Either a semantic or pragmatic strategy is open to a proponent of a content-based view adopting the solution to the Appropriation Worry I argued for here.

Appropriation involves members in the target group working together to subvert the derogatory element of a slur. When appropriation is successful, speakers in the target group can use the slur to express something positive which conveys group solidarity. The account I have argued for captures *Outsider Usage* and *Insider Usage*, while fitting with the general aims and process of social and political movements to reclaim slurs. Next I briefly consider other accounts of slurs that involve indexicality and show how my account differs.

5. *Other Accounts of Slurs with Indexical Elements*

Other theorists have proposed that slurs have an indexical element. Here I briefly discuss three views that involve indexicals and show how they differ from the view for which I have argued.

Schlenker (2007) proposes that slurs and other expressives carry presuppositions that are indexical and attitudinal.[32] For instance, he takes 'honky' to carry the presupposition that the speaker (i.e., agent) of the context of utterance believes that whites are despicable (in the world of the context). The presupposition is indexical and attitudinal as it requires the speaker in the context to have a particular attitude.

There are several differences between the account I have sketched and Schlenker's. First, he focuses on un-appropriated uses of slurs, while I focus on appropriated uses. Second, he take the indexical to be singular rather than plural. I argued above that a plural first-person pronoun better accounts for the Appropriation Worry while being motivated by the general aims of appropriation as solidarity building and group demarcating.

---

[31] There is an apparent tension between adopting a semantic ambiguity strategy to handle appropriation when target-group members use appropriated slurs and a pragmatic strategy to account for "insider" status usage of appropriated slurs. One might argue that if appropriation by a target group is to be accounted for semantically (as content-based theorists for whom the Appropriation Worry is a problem hold) then "insider" status usage should be accounted for semantically as well. I do not take the apparent tension to be decisive. There are after all differences between being friends with women or people racialized as Black and actually being a woman or actually being racialized as Black that could be appealed to in order to motivate a semantic ambiguity strategy on the one hand and a pragmatic strategy on the other. However, the tension might speak in favor of a semantic treatment of "insider" status usage. Thanks to Daniel Wodak and Matthias Jenny for pressing this issue.

[32] He also argues that the presupposition can be shiftable, building off Schlenker (2003) which argues that indexicals are shiftable (i.e., that indexicals can be monstrous). I won't focus on this element of the view.

Thommen gives a brief sketch of a view on which slurs are indexicals. He argues that a slur can only be used in a context in which "the speech participants (the speaker and the hearers) share a certain negative response" to the target group (Thommen 2014: 43). Thommen's view broadens the indexical element of a slur, by taking more than the speaker to be relevant. In doing so, however, he seems to rule out the possibility of a slur being uttered and expressing a derogatory content when some conversational participants do not have negative attitudes towards the target group. For instance, suppose that three people are having a conversation. If two harbor racist attitudes towards Black people and one does not, Thommen's view appears to make a derogatory utterance of a slur against Blacks impossible. I take it that this gets the facts wrong. Moreover, Thommen does not explicitly give a semantics for slurs. So it is not clear whether he wants to include the indexical 'we' or whether he wants to have separate elements that are sensitive to the speaker and hearer(s). In addition, Thommen is focused on unappropriated uses of slurs. He does not aim to address appropriation.

Finally, Kennedy argues that the meaning of a slur can vary "depending upon, among other variables, intonation, the location of the interaction, and the relationship between the speaker and those to whom he is speaking" (Kennedy 2003: 43). Kennedy does not argue for a detailed theory of how a slur's meaning varies, so it is difficult to tell exactly how indexicality is meant to be built into the content of a slur. From the factors listed it sounds like either (a) he takes slurs to have radically indexical contents or that (b) he is including the varied pragmatic effects as parts of a slurs meaning. I am inclined to interpret him as taking the meaning of a slur to include its broad linguistic import—including both conventional lexicalized effects and pragmatic conversational effects. It is certainly right that a slur might have widely varied pragmatic effects depending on, say, whether it was uttered at a basketball game, in a courtroom, or at the United State Holocaust Memorial museum. It is much more controversial, however, whether all these effects should be included in a lexicalized conventional account of the content of a slur. While indexical elements play a role in other treatments of slurs, no one has focused on indexicality as a way to solve the Appropriation Worry and no one has drawn on the semantics of 'we' as I have done here.

6. *Conclusion*

Almost all theorists agree that slurs have derogatory content. We saw that a worry arises for proponents of any content-based approach. If slurs have derogatory content, it is natural to conclude that they gain an additional content when they are appropriated. Yet, the new content can only be expressed by appropriated slurs uttered by members of the target group. The Appropriation Worry stated that content-based views that posit an ambiguity to account for appropriation cannot ac-

count for why only members in the target group can use a appropriated slur to express a non-offensive/positive meaning. To deflate the Appropriation Worry I have argued that the lexical entries for appropriated slurs include a plural first-person indexical. I argued that 'we' is sensitive to one's social position—in particular it is sensitive to the speaker's social group memberships. In general 'we' requires that the speaker actually be a member of the group picked out. The requirement holds regardless of whether the group denoted is just the plurality of the speaker and addressee or a larger group that involves the sharing of gender, racial, ethnic, sexual orientation, occupational, or other features. By considering the aims and purpose of appropriation, social positions, and the way linguistic content can be sensitive to positions in social space, the Appropriation Worry can be solved.

References

Anderson, L. Forthcoming. "Calling, Addressing, and Appropriation." In David Sosa (ed.). *Bad Words*. New York: Routledge.

Anderson, L. and E. Lepore. 2013a. "Slurring Words." *Noûs* 47 (1): 25–48.

Anderson, L. and E. Lepore. 2013b. "What Did You Call Me? Slurs as Prohibited Words." *Analytic Philosophy* 54 (3): 350–363.

Asim, J. 2007. *The N Word: Who Can Say It, Who Shouldn't, and Why*. Houghton Mifflin: Harcourt.

Beaton, M. E. and H. B. Washington. 2015. "Slurs and the Indexical Field: The Pejoration and Reclaiming of Favelado 'Slum-Dweller.'" *Language Sciences* 52: 12–21.

Besson, C. 2010. "Rigidity, Natural Kind Terms, and Metasemantics". In H. Beebee and N. Sabbarton-Leary (eds.). *The Semantics and Metaphysics of Natural Kinds*. New York: Routledge: 25–44.

Bianchi, C. 2014. "Slurs and Appropriation: An Echoic Account." *Journal of Pragmatics* 66: 35–44.

Bolinger, R. J. 2017. "The Pragmatics of Slurs." *Noûs* 51 (3): 439–462.

Butler, J. 1990. *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.

Camp, E. 2013. "Slurring Perspectives." *Analytic Philosophy* 54 (3): 330–349.

Carlson, G. N. 1977. *Reference to Kinds in English*. Ph.D. Dissertation. Amherst: University of Massachusetts.

Cepollaro, B. 2015. "In Defence of a Presuppositional Account of Slurs." *Language Sciences* 52: 36–45.

Cohen, J. and E. Michaelson. 2013. "Indexicality and The Answering Machine Paradox." *Philosophy Compass* 8 (6): 580–592.

Croom, A. M. 2011. "Slurs." *Language Sciences* 33 (3): 343–358.

de Beauvoir, S. 1949/2011. *The Second Sex*. Translated by Constance Borde and Sheila Malovany-Chevallier. New York: Vintage.

Devitt, M. 2005. "Rigid Application." *Philosophical Studies* 125 (2): 139–65.

Elbourne, P. 2008. "Demonstratives as Individual Concepts." *Linguistics and Philosophy* 31 (4): 409–466.

Falkum, I. and Vicente, A. 2015. "Polysemy: Current Perspectives and Approaches." *Lingua* 157: 1–16.

Galinsky, A. D. et al. 2013. "The Reappropriation of Stigmatizing Labels: The Reciprocal Relationship between Power and Self-Labeling." *Psychological Science* 24 (10): 2020–2029.

Grice, H. P. 1975. "Logic and Conversation." In M. Ezcurdia and R. J. Stainton. *The Semantics-Pragmatics Boundary in Philosophy*. Broadview Press: 47–59.

Haslanger, S. 2010. "Gender and Race: (What) Are They? (What) Do We Want Them To Be?" *Noûs* 24 (1): 31–55.

Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. Dissertation. Amherst: University of Massachusetts.

Heim, I. and A. Kratzer. 1998. *Semantics in Generative Grammar*. Malden: Wiley-Blackwell.

Hom, C. 2010. "Pejoratives." *Philosophy Compass* 5 (2): 164–185.

Hom, C. 2008. "The Semantics of Racial Epithets." *The Journal of Philosophy* 105 (8): 416–440.

Hom, C. and R. May. 2013. "Moral and Semantic Innocence." *Analytic Philosophy* 54 (3): 293–313.

Hornsby, J. 2001. "Meaning and Uselessness: How to Think about Derogatory Words." *Midwest Studies in Philosophy* XXV: 128–141.

Hunter, J. 2010. *Presuppositional Indexicals*. Ph.D. Dissertation. University of Texas at Austin.

Jeshion, R. 2013. "Expressivism and the Offensiveness of Slurs." *Philosophical Perspectives* 27 (1): 231–259.

Kaplan, D. 1989. "Demonstratives." In J. Almog, J. Perry, and H. Wettstein (eds.). *Themes from Kaplan*. Oxford: Oxford University Press: 481–563.

Kennedy, R. 2003. *Nigger: The Strange Career of a Troublesome Word*. New York: Vintage.

King, J. C. 2006. "Singular Terms, Reference and Methodology in Semantics." *Philosophical Issues* 16 (1): 141–161.

Kleinman, S. et al. 2009. "Reclaiming Critical Analysis: The Social Harms of 'Bitch.'" *Sociological Analysis* 3 (1): 46–68.

Krifka, M. et al. 1995. "Genericity: An Introduction." In G. N. Carlson and F. J. Pelletier (eds.). *The Generic Book*. Chicago: University of Chicago Press: 1–124.

Langton, R., S. Haslanger, and L. Anderson. 2012. "Language and Race." In G. Russell and D. Graff Fara (eds.). *Routledge Companion to the Philosophy of Language*. New York: Routledge: 753–67.

Leslie, S-J. and A. Lerner. 2016. "Generic Generalizations." *The Stanford Encyclopedia of Philosophy*. Ed. Edward N. Zalta. Metaphysics Research Lab, Stanford University, 2016.

Lycan, W. 2015. "Slurs and Lexical Presumption." *Language Sciences* 52: 3–11.

McConnell-Ginet, S. 2011. *Gender, Sexuality, and Meaning: Linguistic Practice and Politics*. New York: Oxford University Press.

Miščević, N. 2016. "The Fiery Tongue: the Semantics and Pragmatics of Pejoratives." In N. Miščević and J. Perhat (eds.). *A Word Which Bears a Sword: Inquiries into Pejoratives*. Zagreb: Kruzak.

Miyazaki, A. 2004. "Japanese Junior High School Girls' and Boys' First Person Pronoun Use and Their Social World," In S. Okamoto and J. (Shibamoto) Smith (eds.). *Japanese Language, Gender, and Ideology: Cultural Models and Real People.* Oxford: Oxford University Press: 256–274.

Nunberg, G. 1993. "Indexicality and Deixis." *Linguistics and Philosophy* 16 (1): 1–43.

Nunberg, G. Forthcoming. "The Social Life of Slurs." In D. Fogal, D. Harris, and M. Moss (eds.). *New Work on Speech Acts.* Oxford: Oxford University Press.

Partee, B. H. 1989. "Binding Implicit Variables in Quantified Context." *Papers of the Chicago Linguistic Society* 25: 342–365.

Potts, C. 2005. *The Logic of Conventional Implicatures.* Oxford: Oxford University Press.

Potts, C. 2007. "The Centrality of Expressive Indices." *Theoretical Linguistics* 33 (2): 255–268.

Rahman, J. 2012. "The N Word: Its History and Use in the African American Community." *Journal of English Linguistics* 40 (2): 137–171.

Recanati, F. 1993. *Direct Reference: From Language to Thought.* London: Blackwell.

Richard, M. 2008. *When Truth Gives Out.* Oxford / New York: Oxford University Press.

Saka, P. 2007. *How to Think about Meaning.* Dordrecht: Springer.

Schlenker, P. 2007. "Expressive Presuppositions." *Theoretical Linguistics* 33 (2): 237–245.

Schlenker, P. 2003. "A Plea for Monsters." *Linguistics and Philosophy* 26: 29–120.

Sennet, A. and D. Copp. 2015. "What Kind of a Mistake Is It to Use a Slur?" *Philosophical Studies* 172 (4): 1079–1104.

Thommen, T. 2014. *Response-Dependence and the Semantics of Slurs.* MA Thesis, Institut Jean Nicod.

Wechsler, S. 2010. "What 'you' and 'I' mean to Each Other: Person Indexicals, Self-Ascription, and Theory of Mind." *Language* 86 (2): 332–365.

Whiting, D. 2013. "It's Not What You Said, It's the Way You Said It: Slurs and Conventional Implicatures." *Analytic Philosophy* 54 (3): 364–377.

Williamson, T. 2010. "The Use of Pejoratives." In D. Whiting (ed.). *The Later Wittgenstein on Language.* Basingstoke: Palgrave-Macmillan.

Wodak, D. and S-J. Leslie. Forthcoming. "The Mark of the Plural: Generic Generalizations and Race." In P.C. Taylor, L.M. Alcoff and L. Anderson (eds.). *Routledge Companion to Philosophy of Race.*

Zwicky, A. M. and J. M. Sadock. 1975. "Ambiguity Tests and How to Fail Them." In J. P. Kimball (ed.). *Syntax and Semantics*, 4. New York: Academic Press: 1–36.

Let's Not Worry about the Reclamation Worry

BIANCA CEPOLLARO*
IFILNOVA Institute of Philosophy,
New University of Lisbon, Lisbon, Portugal

In this paper, I discuss the Reclamation Worry (RW), raised by Anderson and Lepore 2013 and addressed by Ritchie (2017) concerning the appropriation of slurs. I argue that Ritchie's way to solve the RW is not adequate and I show why such an apparent worry is not actually problematic and should not lead us to postulate a rich complex semantics for reclaimed slurs. To this end, after illustrating the phenomenon of appropriation of slurs, I introduce the Reclamation Worry (section 2). In section 3, I argue that Richie's complex proposal is not needed to explain the phenomenon. To show that, I compare the case of reclaimed and non-reclaimed slurs to the case of polysemic personal pronouns featuring, among others, in many Romance languages. In section 4 I introduce the notion of 'authoritativeness' that I take to be crucial to account for reclamation. In section 5, I focus on particular cases (the "outsider" cases) that support my claims and speak against the parsimony of the indexical account. Finally, I conclude with a methodological remark about the ways in which the debate on appropriation has developed in the literature (section 6).

Keywords: Reclamation worry, appropriation, ambiguity, slurs, polysemy.

1. *Introduction*

In this paper, I discuss a worry raised by Anderson and Lepore (2013) and addressed by Ritchie (2017) concerning the appropriation of slurs, namely the Reclamation Worry (RW). My aim is to show that despite appearances the Reclamation Worry is not worrisome and could be addressed by simply relying on contextual meaning determination: no indexical account *à la* Ritchie is needed. To this end, I present a more parsimonious answer to the apparent problem raised by the RW. Finally, I introduce and discuss the notion of 'authoritativeness' and conclude with a methodological remark about the ways in which the debate on appropriation has developed in the literature.

Let us start from defining Reclamation. Reclamation (or 'appropriation', I use the terms interchangeably) is the phenomenon for which, under certain conditions, speakers can use a slur in such a way that the slur is not derogatory nor offensive anymore in those contexts; on the contrary, appropriated slurs are used to express solidarity and underline intimacy. Typically, in-groups can use the slur targeting their own group in such a reclaimed way. However, over time, non-derogatory uses can become available for out-groups, too. This is for example what happened to the term 'gay', that used to be derogatory and nowadays it is neutral, after a process of appropriation (see Brontsema 2004). This suggests that reclamation is able to challenge the derogatory potential of slurs and, in time, even delete it. Such a process is on-going for other terms that still have a derogatory use (such as 'queer') or terms that can typically be used non-derogatorily by in-groups only (such as 'nigger').

Reclamation is a very problematic topic in the growing literature[1] about pejoratives partly because it constitutes an example of meaning change: thanks to reclamation, slurs can be used, under adequate conditions, in a positive way that differs "dramatically" from their original derogatory meaning (see Potts 2007: 266). One can account for such a change of meaning in (at least) two different ways: by defending an echoic account or by defending a polysemic account. The echoic explanation consists in interpreting the reclaimed positive use of slurs as non-literal derivative uses of language (see Bianchi 2014, Miščević and Perhat 2016: 140, Cepollaro 2017). Slurs have derogatory *literal* meaning also in reclaimed uses, but speakers can use it in a positive way thanks to irony. Reclamation is analyzed by these authors as an instance of ironic use of language (which is, in relevance theoretic terms, an *echoic* use of language: Sperber and Wilson 1986, Wilson and Sperber 2012). An alternative way to understand the change of meaning stemming from appropriation is to claim that slurs, under the right circumstances, acquire a new non-pejorative *literal* meaning (see i.a. Miščević 2011, Ritchie 2017). Such a process results in the ambiguity

---

[1] See i.a. Potts (2007: 266), Hom (2008: 428, 438), Richard (2008: 16), Saka (2007: 146-147), Jeshion (2013: 250-253), Whiting (2013: 370), all quoted in Ritchie (2017).

between two lexical items: the old derogatory slur and the reclaimed positive ex-slur. The Reclamation Worry, to which I now turn, especially targets such a polysemic account of reclamation. In this paper, I leave aside the echoic account, as my main (and more modest) aim here is just to establish whether the Reclamation Worry (i) is indeed worrisome and (ii) should be taken as supporting an indexical semantics for reclaimed uses of slurs.

2. *The Reclamation Worry*

Ritchie (2017) phrases the Reclamation Worry, already voiced by Anderson and Lepore,[2] in the following way:

> Content-based views that posit an ambiguity to account for reclamation cannot account for why only members in the target group can use a reclaimed slur to express a non-offensive/positive meaning. (Richie 2017: 157)

It looks problematic – or at least challenging – to explain why, given two meanings that an expression can have, one particular meaning is available to a certain subset of speakers, but not to others. There seems to be a contrast, Ritchie argues, between standard cases of ambiguity/polysemy and the behavior of appropriated and non-appropriated uses of slurs. In standard cases of polysemy, all the meanings that an expression can have are potentially available to *any* speaker. Consider standard instances of polysemy, for example, 'mouse'. In principle, anyone can access any of the two meanings of 'mouse', as anyone can use the term to refer either to the mammal or to the electronic device. In other words, there are no restrictions with respect to *who* can felicitously access each meaning. On the other hand, not every speaker can felicitously use slurs as reclaimed.

Since scholars got interested in slurs and pejoratives, the phenomenon of appropriation has caught the attention of many. The question as to how and why reclaimed uses of slurs are possible for certain people and not for others challenged philosophers and linguists, but it also gave rise to a debate outside academia: there are many ways in which tv-series, movies and even cartoons assess the issue of who can and who cannot use slurs in positive ways.[3]

---

[2] "Ambiguity fails to explain why non-members cannot utilize a second sense [that is, the non-pejorative sense]. If it were just a matter of distinct meanings, why can't any speaker opt to use a slur non-offensively? (…) Ambiguity is useless here" (Anderson and Lepore 2013a: 42).

[3] Here are some interesting examples. In the tv-series *Treme* (2010), Season 1 episode 5 titled 'Shame shame shame', Davies, a white guy, gets punched after using the word 'nigger' in a bar attended by black people, even though he was identifying himself with the black community. An interesting fact is that Davies is explicitly *quoting* another character, Antoine Baptiste, who is in fact black ("I can only quote Antoine Baptiste: New Orleans niggers will fuck up a wet dream. Media freak the fuck out. The cops looking for any excuse to clamp down"). The explicit quotation does not save him from getting punched, nor does the fact that he identifies with the black community of New Orleans.

In her paper, Ritchie offers a solution to the Reclamation Worry that accounts for the allegedly peculiar ambiguity between reclaimed and non-reclaimed slurs by claiming that the lexical meaning of reclaimed slurs features an occurrence of the plural first-person indexical pronoun 'we'. A reclaimed use of 'bitch', for example, conveys for Ritchie something along the lines of "*We* women are laudable for being women".[4] Only in-groups (women, in this case) can access the reclaimed slur because they are the only ones who can *felicitously* use the indexical 'we'. When an out-group (a man, in this case) tries to use 'bitch' in a positive way, he fails to convey a positive content such as "*We* women are laudable for being women", because he cannot felicitously use the indexical 'we'.

In other words, Ritchie grants that the RW constitutes a challenge for content-based accounts of slurs that explain reclamation in terms of ambiguity, as appropriation gives rise to a particular case of ambiguity that diverges from standard instances of in that, because of the indexical 'we', only certain speakers can felicitously access the reclaimed meanings. The challenge for Ritchie's indexical account is to explain how out-groups can felicitously use slurs in a positive or non-derogatory way: I come back to this issue in greater detail in section 5.

In what follows, I show that despite appearances the Reclamation Worry is not worrisome and that an indexical account of reclaimed slurs is not needed. I will not *refute* Ritchie's proposal: rather, I will show that (i) no indexical pronoun in the reclaimed use of slurs is to be postulated to satisfactorily answer the Reclamation Worry (section 3) and that (ii) all the ingredients that are needed to answer the RW *without* postulating hidden indexicals are already employed by Ritchie to explain the outsider cases, namely out-groups using slurs non-derogatorily (section 4). I conclude that more parsimonious responses to the RW should be preferred and that the indexical explanation of Reclamation needs further justification to be defended.

Before discussing my own solution to the RW, let me make a brief remark about the theories that are allegedly challenged by such a worry. The authors who discussed the Reclamation Worry, including Ritchie, focus on the so-called 'content-based' account of slurs, that is, the theories according to which slurs lexically encode some kind of de-

---

In the 2016 cartoon Zootopia, a bunny police-officer is called 'cute' by her cheetah-colleague; she tells him: "Ooh, uh, you probably didn't know, but a bunny can call another bunny 'cute', but when other animals do it, it's a little...". The cheetah police-officer is mortified.

In the tv-series *Atlanta* (2016), Season 1 episode 04 titled 'The Streisand Effect', a character named Zan goes to Alfred, a black rapper, and calls him "My nigga". Alfred's answer is "Are you even black?". Afterwards the characters discuss and make hypotheses about Zan's mysterious ethnic origins which are taken to be crucial in order to establish whether his use of 'nigga' was legitimate or not.

[4] Note that Ritchie is neutral with respect to which particular content-based account captures best the semantics and the pragmatics of slurs. Such a proposal should therefore be taken to be compatible with possibly any content-based view.

rogatory content (however such a lexical component could be analyzed: truth conditions, conventional implicatures, presuppositions, etc.). This label is meant to distinguish this first type of theories from those that challenge the very idea that a thick and rich semantics for slurs is needed. For such accounts, that we can call content-less (or "deflationary") (Anderson and Lepore 2013a), the derogatory content associated with slurs is *not* part of their encoded or lexical meaning. They account for the pejorative power of slurs by relying on various mechanisms: Anderson and Lepore (2013a, 2013b) talk about taboo effects, edicts and language prohibitions, Bolinger (2015) relies on contrastive preferences, Nunberg (forthcoming) introduces the notion of manner ventriloquistic implicatures, Rappaport (ms) analyzes the pejorative content of slurs in terms of the relevance-theoretic notion of 'showing'. *Prima facie*, the phenomenon of appropriation is less problematic for such accounts, as they are not committed to the claim that slurs lexically encode derogatory content (and therefore they do not have to posit polysemy to explain appropriation). However, it is not entirely correct to conclude that they are not challenged by the meaning change that we observe in appropriation, as, if these theories need to explain how slurs are *systematically* associated with derogatory contents, then they also need to explain how and why such an association can fail to hold in reclamation contexts. So, *if* the Reclamation Worry was posing a problem for content-based account of slurs, then it would potentially pose a challenge for all theories of slurs. In what follows, I argue that this is not the case.

3. *Ambiguity and social meaning*

As we said, the Reclamation Worry was originally voiced by Anderson and Lepore as a potential argument against the content-based theories, even though the authors already suggest that in order to save the account, content-based theorists could add an extra story to explain why the non-derogatory use of slurs is only accessible to a certain subset of speakers. This 'extra story' is exactly what Ritchie aims to add, by arguing that the lexical meaning of reclaimed slurs involves the occurrence of a plural first-person pronoun that imposes constraints on who can use the reclaimed slur: for instance a reclaimed use of 'bitch' conveys a content along the lines of "*We* women are laudable for being women". In other words, Ritchie endorses the idea that the ambiguity between reclaimed and non-reclaimed uses of slurs deeply differs from standard instances of ambiguity and provides a story to explain why appropriation gives rise to such a unique case of ambiguity.

In this section I present an alternative and simpler solution. My claim is that the ambiguity between reclaimed and non-reclaimed slurs is not particular nor problematic and I try to show that other instances of ambiguity display similar properties: the context, together with speaker's intentions, provides constraints on who can felicitously

access a certain meaning of a polysemic term in each situation. As a matter of fact, (i) contexts typically include information about the relations among the speakers as well as information about the beliefs and attitudes of the participants to the conversation and (ii) such information can drive the interpretation of expressions and utterances.

Let us now turn to the instance of ambiguity that I take to prove the non-exceptionalness of reclamation. Consider for instance the use of '*tu*' and '*vous*' in French or '*tu*' and '*lei*' in Italian. The '*vous*' and '*lei*' forms are the formal pronouns used to formally address the interlocutor in French and Italian respectively. More importantly, both of them are identical to *another* personal pronoun that is neutral with respect to the parameter of formality: in French, '*vous*' can be either the formal singular second-person pronoun or the formal/informal plural second-person pronoun; in Italian '*lei*' can be either the formal singular second-person pronoun or the formal/informal singular third-person pronoun. Suppose that A addresses B in Italian and says:

(1) Lei gradisce del tè?
    (a) Would *you* [formal] like some tea?
    (b) Would *she* [informal/formal] like some tea?

As it is, (1) is *ambiguous* between (1a) "Would *you* like some tea?" (formal) and (1 b) "Would *she* like some tea?" (formal or informal), because the formal second-person pronoun '*lei*' is identical to the singular third-person pronoun '*lei*' (formal or informal).

Given the social information provided by the context, the addressee will typically access one interpretation or the other and understand whether the term at stake is a second or third-person pronoun. Such information about the social relations of the speaker does not only drive the interpretation from the point of view of the addressee: it also constrains which meaning of '*lei*' ('you' or 'she') each speaker can felicitously express. We could describe the situation as follows: only those speakers who are in an *informal* relation to the addressee can felicitously use (1) and be taken to say "Would *she* like some tea?"; and only the speakers who are in a *formal* relation to the addressee can felicitously use (1) to mean (and be taken to mean) "Would *you* like some tea?". Note also that for the use of (1) (as meaning (b)) to be felicitous, it is *not* enough that the speaker *thinks* that she is in an informal relation to her addressee; also her addressee has to recognize her as being in an informal relation with him. The same goes if the speaker intends to use (1) as meaning (a): for such an utterance to be felicitous, it is not enough that the speaker *thinks* that she is in a formal relation to her addressee; also her addressee has to recognize her as being in a formal relation with him. The same goes for the French and Spanish equivalent cases of polysemic pronouns.

The '*lei*' example suggests that the Reclamation Worry raised by Anderson and Lepore might be just an apparent worry. The ambiguity between formal and informal personal pronouns shows that the am-

biguity between negative and positive senses of slurs is not the only case where social information about the speaker and about her relation to the others can affect the way in which a polysemic term is both used and interpreted. The clues that help us decide in which sense a polysemic term is used are provided by the context (broadly construed) and by the intentions of the speakers. If that is correct, then the Reclamation Worry should not be seen as an objection to content-based theories, as Anderson and Lepore suggested, nor as evidence in favor of an indexical account of the lexical content of reclaimed slurs, as Ritchie suggests. The ambiguity between reclaimed and non-reclaimed slurs would be just another instance of the following phenomenon: socially important information determines and constrains the senses in which terms can be used by speakers and interpreted by hearers.

As we shall see in greater detail in Section 5, the indexical account proposed by Ritchie has to strongly rely on contextual factors such as speaker intentions and the like in order to explain the 'outsider' cases, namely – in her terminology, that I adopt – the cases where out-groups can use slurs non-derogatorily. According to a naïve version of the indexical account of reclaimed slurs, a felicitous outsider use of a reclaimed slur could not be, due to an infelicitous use of the indexical 'we'. We shall see that in order to fix such a problem, Ritchie employs similar resources to those we are invoking here to explain the ambiguity between reclaimed and non-reclaimed slurs. I will conclude that a more parsimonious answer to the RW should be preferred.

4. *Authoritativeness*

It is surely a useful simplification to hold that what is typical of appropriation is that it starts and spreads among in-groups. However, there are good reasons *not* to frame the Reclamation Worry in terms of appropriated uses being accessible for in-groups only and offensive uses being accessible for out-groups and possibly in-groups. If the shift or extension from in-groups to out-groups was not available, it would not be possible to explain outsider cases, nor how reclamation can finally lead to slurring terms losing their derogatory power for good. If that is correct, what makes a non-pejorative use of slurs possible in general does not amount to the category(s) to which the speaker belongs, but rather, how likely it is for the speaker to be taken as *genuinely* and felicitously expressing a positive attitude towards the target class. In other words, what is at stake is whether the audience accepts the speaker's intention of dissociating from the negative use of slurs and subverting such a use, not necessarily whether the speaker belongs to the target group or not. Is then the in-group/out-group description of appropriation just wrong? Of course not. To be an in-group is the safest way (or one of the safest ways) for a speaker to be accepted as genuinely[5] (i) expressing

---

[5] One reason to stress that the attitude has to be recognized as 'genuine' is that people who do not undergo certain kind of discriminations themselves are not always

a dissociative attitude with respect to bigot beliefs typically conveyed by slurs and (ii) communicating a positive attitude[6] towards the target group. However, as many cases show, this does not need to be the case. The mere fact that reclaimed uses *can* be open to out-groups signals that the right direction to go in phrasing the Reclamation Worry is not in terms of in-groups and out-groups but in terms of 'believability' or 'authoritativeness' of the subject. How much a subject is taken to be authoritative depends on many complex factors that vary on a case-by-case basis (hence, the discussion inside and outside academia concerning who can use reclaimed slurs; see footnote 3): for sure it is important with which groups she can identify with, but also what kind of experiences she had undergo, how clear her beliefs and her stance towards bigotry are and so on. My claim is that being an in-group is just *one* way to be very authoritative and I present a scenario that I take to support my claim. Take three men, John, Peter and Bob. They are gay. John is an activist and spent his entire life fighting homophobia. Peter, on the other hand, never felt like telling anyone that he is gay, except for very few people. He is very discreet about it and never participates to LGBTQ+ pride events. He never engages in discussions about gay marriage nor anything related to LGBTQ+ rights. Suppose Bob is a close friend of both and knows they are gay. One day, on different occasions, Bob hears them talking about a common friend being 'a fag'. My intuition is that John is somehow more entitled than Peter to use the slur 'fag' in a positive way. In particular, I would expect Bob to have no doubts in interpreting John's use of the slur as reclaimed and to feel uneasy or dubious about Peter's use of the term. If what allows speakers to access positive uses of slurs was 'just' their belonging to the target group, and if we grant that the group targeted by the slur 'fag' is gay men, then we cannot account for the intuition that John's reclaimed use of a homophobic slur is more acceptable than Peter's use. The only way to save the in-groupness account of reclamation would be either (i) to say that Peter should count as less of an in-group than John, which does not make much sense if the target of the homophobic

---

taken to be entitled to express solidarity attitudes nor to use appropriation. Think about the fact that while right now, as Ritchie notices, it is nearly impossible for men in general to use 'bitch' in a reclaimed way, gay men are sometimes (sometimes!) allowed to do so.

[6] An interesting issue is whether the attitude expressed in reclamation contexts has to be positive or just non-negative. As I see it, in the first phases of reclamation, the attitude expressed by the speaker/the content conveyed by the term (this depends on what theory of appropriation one favours) has to be positive for a certain occurrence of a slur to count as – and to be taken as – reclaimed. However, since appropriation can lead to a point where the slur loses its derogatory power for good and becomes a neutral term (think of the case of 'gay') rather than a positive word, then it's plausible to think of reclaimed uses of slurs such that the speaker does not convey any positive content, she just fails to convey negative ones. I'd like to thank Erich Rast for pushing me on this point.

slur 'fag' are gay men, or (ii) to say that the group targeted by 'fag' is not gay men, but gay activists, which is also quite implausible.

A more promising way to explain why Peter could be perceived as less entitled than the activist John to use 'fag' in a positive way is in terms of authoritativeness. It is easier for Bob to recognize John's anti-homophobic intentions and attitudes, while he would harbor doubts about whether or not Peter's dissociative attitude with respect to homophobia and Peter's solidarity attitude with respect to the LGBTQ+ are truly genuine. And, again, this has nothing to do with Peter's being a member of the target class or not. It is about the *authoritativeness* that Peter is granted in a group.

The more a slur gets reclaimed, the less authoritativeness is needed to felicitously access the non-derogatory meaning: the less problematic a term becomes, the less is needed to recognize the speaker's intentions as felicitously communicating non-derogatory attitudes.

We cannot account for the John-Peter case intuitions if we stick to the in-group/out-group schema, nor can we account for the outsider cases: we need to think in terms of authoritativeness, legitimacy and believability in order to account for the criteria on the basis of which speakers can or cannot access reclaimed uses of slurs.

5. *In- and out-groupness: the outsider cases*

In Section 3, I argued that the Reclamation Worry does not arise at all, since it is not surprising that in cases of polysemy or ambiguity, complex social information is required for the speaker to use the ambiguous term with a certain meaning and for the hearer to interpret it correctly. In doing so, I claimed that the indexical account of appropriated slurs should not be invoked as a solution to the Reclamation Worry, since the worry does not arise in the first place. Moreover, in Section 4, I claimed that the main criterion with which certain uses of slurs are interpreted as reclaimed seems to deal with recognizing and accepting speakers' intentions on the basis of the authoritativeness they are granted, rather than just considering whether speakers belong or not to a group.

In this section, I come back to the outsider cases to show that they speak in favor of my claim and reveal some lack of parsimony on the side of the indexical account of reclaimed slurs.

The outsider cases are, as we said, those cases where out-groups use slurs non-derogatorily. The reclaimed use of slurs like 'queer' is a good and well-known example, since out-groups can felicitously use a slur in a positive or neutral way[7] (think of expressions such as Queer Studies, Queer Tango, Queer Film Festival, the acronym LGBTQ+, etc). But there are many other cases where the possibility for the outsider's re-

---

[7] Ritchie excludes from her investigation cases such as 'queer', where the process of appropriation led to a point where a speaker clearly does not need to identify with the target class in order to use the slur in a reclaimed way. However, I think it is crucial for a satisfactory account of reclamation to be able to explain this phase, too.

claimed use to be felicitous or not depends very much on the context: under the adequate circumstances, an out-group can use a slur in a positive way, but it does not only depend on her intentions. It is crucial for the indexical account to explain how these uses are even possible. As a matter of fact, it looks *prima facie* very hard for the indexical account of reclaimed slurs to explain how an out-group could felicitously use a slur in a positive way: if reclaimed slurs involve the occurrence of an indexical 'we', when an out-group uses a 'we', it should be infelicitous. Ritchie addresses the problem by introducing the notion of 'insider' status. The outsider cases are not infelicitous because under certain circumstances out-groups can get something like an insider status, they can count as belonging to a 'we' even though they do not belong to the target group. The 'insider' status makes their use of the indexical 'we' felicitous despite the out-groupness of the speaker.

I would resist this move for two reasons. The first one is that such an explanation does not account for all the data of reclaimed slurs, but only for a subset. In fact, while it could be an adequate description of some outsider uses, speakers need not always perceive (and be taken as perceiving) themselves and the target class as a 'we' in order to use reclaimed slurs. For instance, a scholar using the term 'queer' in talking about 'Queer Studies' is not necessarily identifying herself with the target class (she can of course; but she does not *need* to). The outsider cases actually show that what is at stake is not whether a speaker belongs or not to a category, nor whether she can talk about the target class using a plural first-person pronoun. The beliefs and intentions of each one in each context, together with the way in which such beliefs and intentions are recognized and accepted, seem to constitute what makes a use of a slur reclaimed or not. Then, again, as it was said in Section 4, belonging to a group is a strong contextual clue of what the speaker's intentions and attitudes are, but it is not all there is.

Moreover, the indexical account ends up relying on notions such as intention attribution and recognition, as well as something close to authoritativeness (namely the 'insider' status), in order to explain the outsider cases; but once one invokes these notions to explain the outsider uses, these notions turn out to be just enough to account for *all* reclaimed uses. In other words, there is no need to postulate hidden indexicals for reclaimed uses of slurs, if the outsider cases already require us to rely on intention interpretation, and some relevant social relations involving the speaker (authoritativeness, in my terminology). My conclusion is that a more parsimonious response to the RW should be preferred to the indexical account and that the indexical explanation of Reclamation would need further justification to be defended.

6. *Conclusion: A brief methodological remark*

The debate on appropriation started when philosophers of language and linguists became interested in slurs and pejorative language. However, it is easy to observe instances of valence reversal that we might call 'reclamation' for any kind of evaluative term. Slurs are not the only evaluative words that used to be negative and that were used positively until the point where the term lost its pejorative power for good. I consider just two famous examples from the history of poetry and painting. Let us start from the most ancient one. In a letter to Atticus,[8] Cicero calls a group of poets '*neoteroi*' (from the greek 'νεώτεροι', 'the newer ones'), disregarding this avant-garde interested in introducing in poetry new stylistic features and new themes. Two centuries after, a group of poets called themselves 'the novel poets', after the neoteroi. The term lost its derogatory power, even if it was initially created and used as derogatory by Cicero. First, it was used positively by those later poets who were very much inspired by the group that Cicero scornfully called '*neoteroi*' and nowadays '*neoteroi*' is how the scientific community neutrally calls these very much appreciated poets. Another (possibly more famous) example concerns the impressionists. The French term '*impressioniste*' was disdainfully introduced by a critic, Louis Leroy, in the journal *Le Charivari* in 1874.[9] Not long after, the term was used again, but in a positive sense, by Jules-Antoine Castagnary in his paper "Exposition du boulevard des Capucines. Les impressionnistes", published on *Le Siècle*.[10] Nowadays, we still call these artists 'impressionists' with no trace of disdain. Besides these examples, we can find many more instances of appropriation from very different fields that have nothing to do with slurs as they are commonly understood.[11] Just to mention a few, the English adjective 'terrific', coming from the Latin adjective 'terrificus' (frightening), underwent a process of valence change. Until 1880s, it meant – just like in Latin and contemporary Italian – 'horrible' or 'frightening'. From around the 1880s, it starts meaning 'excellent', which is its standard meaning today in English.[12] Or take the adjective '*bárbaro*' in Spanish: it used to be exclusively negative (mean-

---

[8] From the *Epistulae ad Atticum*, Book 7, Letter 2 (year 50 B.C.). Latin text available online: http://perseus.uchicago.edu/perseus-cgi/citequery3.pl?dbname=PerseusLatinTexts&query=Cic.%20Att.%207.2&getid=0.

[9] Leroy, Louis (1874), L'Exposition des impressionnistes, Le Charivari, April 25th 1874.

[10] Castagnary, Jules-Antoine (1874), Exposition du boulevard des Capucines. Les impressionnistes, Le Siècle April 29th, pp. 1-10.

[11] Nunberg (forthcoming) claims that that of 'slur' is a fairly recent notion. I do not agree with Nunberg on this point, but even if he was right, the 2000 year old case of 'neoteroi' shows that appropriation is not a recent nor slurs-only-related phenomenon.

[12] http://www.etymonline.com/index.php?allowed_in_frame=0&search=terrific&searchmode=none.

ing 'horrible') and nowadays it can *also* be used positively ('fantastic').[13]

The reason why I mention these apparently unrelated cases is to make a methodological point: appropriation should be seen as a general phenomenon that is *not* restricted to slurs.[14] The reclamation of slurs is just one instance of valence change. I grant that the possibility of reversing the valence of an evaluative term becomes particularly crucial and politically precious when it comes to socially loaded terms like slurs, but this does not mean that we ought to study appropriation as a characterizing feature of slurs *only*. When investigating the appropriation of slurs, it is important to bear in mind that it is a general mechanism, not only related to slurs. In order to provide a satisfactory account of appropriation, a much broader stance is required.

References

Anderson, L. and Lepore, E. 2013a. "Slurring words." *Nous* 47 (1): 25–48.

Anderson, L. and Lepore, E. 2013b. "What did you call me? Slurs as prohibited words." *Analytic Philosophy* 54 (1): 350–363.

Bianchi, C. 2014. "Slurs and appropriation: An echoic account." *Journal of Pragmatics* 66: 35–44.

Bolinger, R. J. 2015. "The Pragmatics of Slurs." *Noûs* 49 (1): 1–24.

Brontsema, R. 2004. "A queer revolution: reconceptualizing the debate over linguistic reclamation." *Colorado Research in Linguistics* 17: 1–17.

Cepollaro, B. 2017. "When evaluation changes. An echoic account of appropriation and variability." *Journal of Pragmatics* 117: 29–40.

Hom, C. 2008. "The semantics of racial epithets." *Journal of Philosophy* 105: 416–440.

Jeshion, R. 2013. "Expressivism and the Offensiveness of Slurs." *Philosophical Perspectives* 27 (1): 231–259.

Jeshion, R. 2017. *Pride and Prejudiced*. Handout for the Evaluatives and Expressives Workshop (Milan).

Miščević, N. 2011. "Slurs & Thick Concepts. Is the New Expressivism Tenable?" *Croatian Journal of Philosophy* 11 (2): 159–182.

Miščević, N. and Perhat, J. 2016. *A word which bears a sword*. Zagreb: KruZak.

Nunberg, G. Forthcoming. "The Social Life of Slurs." In D. Fogal, D. Harris, and M. Moss (eds.). *New Work on Speech Act*. Oxford: Oxford University Press.

Rappaport, J. (ms), Communicating with Slurs.

Richard, M. 2008. *When Truth Gives Out*. Oxford: Oxford University Press.

Ritchie, K. 2017. "Social Identity, Indexicality, and the Appropriation of Slurs." *Croatian Journal of Philosophy* 17 (2): 155–180.

---

[13] Real Academia Española, http://dle.rae.es/?id=52DLHf1.

[14] Jeshion (2017) argues that not only appropriation does not only concern slurs and labels: it does not only concern language. You can have appropriation in any domain, including non-linguistic ones such as fashion. Moreover, Jeshion (2017) offers a list of ways in which meaning change can happen: I do not address the varieties of meaning change here, but I am sympathetic to Jeshion's views.

Potts, C. 2007. "The Centrality of Expressive Indices." *Theoretical Linguistics* 33 (2): 255–268.

Saka, P. 2007. *How to Think about Meaning.* Dordrecht: Springer.

Sperber, D. and Wilson, D. 1986. *Relevance: Communication and Cognition.* Cambridge: Harvard University Press.

Whiting, D. 2013. "It's Not What You Said, It's the Way You Said It: Slurs and Conventional Implicatures." *Analytic Philosophy* 54 (3): 364–377.

Wilson, D. and Sperber, D. 2012. *Meaning and Relevance.* Cambridge: Cambridge University Press.

The Myth of Embodied Metaphor

NIKOLA KOMPA*
Institute of Philosophy,
University of Osnabrück, Osnabrück, Germany

According to a traditionally influential idea metaphors have mostly ornamental value. Current research, on the other hand, stresses the cognitive purposes metaphors serve. According to the Conceptual Theory of Metaphor (CTM, for short), e.g., expressions are commonly used metaphorically in order to conceptualize abstract and mental phenomena. More specifically, proponents of CTM claim that abstract terms are understood by means of metaphors and that metaphor comprehension, in turn, is embodied. In this paper, I will argue that CTM fails on both counts.

Keywords: Conceptual theory of metaphor, embodied metaphor, abstract terms, simulation.

1. Introduction

According to a traditionally influential idea, metaphors have mostly ornamental value. Yet current research in philosophy, linguistics and psychology points in a different direction and stresses the cognitive functions metaphors might serve. According to a rather popular, contemporary account of metaphor, the so-called Conceptual Theory of Metaphor, expressions are commonly used metaphorically in order to conceptualize abstract and mental phenomena. Metaphor comprehension, in turn, is said to be embodied. In the background is a family of

theories that are united by the claim that language comprehension in general is embodied.

In this paper, I will first sketch the basic idea of embodied approaches to language comprehension and examine some of the evidence that has been brought forth in favor of the claim that language comprehension in general is embodied. Then the problem of abstract terms will we raised; abstract terms make trouble for embodied approaches as sensorimotor information doesn't seem to be required for understanding these terms. The remainder of the paper will be devoted to the topic of metaphor. From an embodied point of view, metaphors are not only an interesting phenomenon in its own right; they also promise to help explain how we come to understand abstract terms. Consequently, I will address the questions of whether we understand abstract terms by means of metaphors and whether metaphor comprehension itself is embodied. Eventually, I will argue that both questions have to be answered in the negative.

2. *Embodied Approaches to Language Comprehension*

Within the paradigm of Situated Cognition,[1] which spans a rather wide and varied research area, the claim that cognition is embodied has gained some prominence (cf., e.g., Shapiro 2011, 2014). My concern will be with (so-called) embodied approaches to language comprehension. The following quote by Jerome Feldman and Srinivas Narayanan epitomizes the central idea

> that all understanding involves simulation or enacting the appropriate embodied experience. When asked to grasp, we enact it. When hearing or reading about grasping, we simulate grasping (…). (Feldman and Narayanan 2004: 389)

Moreover, the enactment or simulation in question

> is not grounded in previous sensorimotoric experiences of a generic sort, but instead invokes rather specific sensorimotoric experiences. (Scorolli 2014: 127)

This is also put by saying that modality-specific representations, as opposed to amodal, abstract representations, are evoked in language comprehension. These representations are similar to those we form when we directly experience our environment (cf. Barsalou 1999). They are located in regions of the brain devoted to action and perception. "Cognition is inherently perceptual, sharing systems with perception at both the cognitive and neural levels."—Lawrence Barsalou claims (Barsalou 1999: 577); and language comprehension is a case in point. So the idea, basically, is that in order to understand a linguistic expression one has to simulate the corresponding experience. When I hear the word 'grasp'

---

[1] Here is a popular way of partitioning the field: "According to our usage, then, situated cognition is the genus, and embodied, enactive, embedded, distributed cognition and their ilk are species" (Robbins and Aydede 2009b: 3).

I simulate (reenact) the action of grasping. And since I have grasped before, I will be successful and come to understand the word in question. Simulation, in turn, requires activation in sensory and motor (as well as affective) regions of the brain because in simulating a particular experience we exhibit roughly the same pattern of neural activity that accompanied the initial experience. Language comprehension crucially involves recruitment of the sensory-motor system.

Yet why speak of embodiment here? As Philip Robbins and Murat Aydede point out

> ... not all forms of embodiment involve bodily dependence in the strict and literal sense. Indeed, most current research on embodiment focuses on the idea that cognition depends on the sensorimotor brain, with or without direct bodily involvement. (Robbins and Aydede 2009b: 5)

There are other notions of embodiment; and they all need sorting out. For now, let the Embodiment Claim (EC, for short) be the following claim:

EC_{sim}: Simulation is necessary for language comprehension.

Those who march under the banner all agree that EC is "supported by a growing body of evidence…" (Kaschak et al. 2014: 118)

3. *Evidence*

Evidence is provided by studies that investigate whether during language processing there is activation in regions of the brain devoted to action and perception. There are different types of studies; neuroimaging and behavioral studies, e.g. (for an overview, cf. Kaschak et al. 2014). Here is a short selection of some extensively discussed studies.

First, there are fMRI-studies that show that processing verbs, which denote actions performed by hand (*pick*, *grasp*), foot (*kick*) or mouth (*lick*) elicits activation in the motor (and premotor) cortex (in a somatotopically organized manner). Various studies by Friedmann Pulvermüller, Olaf Hauk, Lisa Aziz-Zadeh and others found a strong congruence between those areas in the brain that are activated during the observation of actions performed by hand, mouth and foot and those areas activated during processing linguistic phrases relating to hand, mouth and foot (Hauk et al. 2004; Pulvermüller 2005, Aziz-Zadeh et al. 2006). The same region is activated when seeing someone grasp a handle and when hearing the sentence "She grasps the handle", for example.

Secondly, there are various behavioral studies. In a so-called action-sentence congruity task, Arthur Glenberg and Michael Kaschak (2002) asked participants to judge whether sentences are sensible or not.[2] The (sensible) sentences in question imply either movement away or move-

---

[2] As Fred Adams points out, it is not clear what Glenberg and Kaschak mean by "sensible". It cannot mean "meaningful", because in the case of sentences which are not sensible (such as "Hang the coat on the upright cup") "it is because one *knows* what these sentences *mean* that one can *tell* that they are FALSE …" (Adam 2010: 623)

ment towards the body ("He opens the drawer"/"He closes the drawer"). Answering the sensibility question also requires either a movement away from or towards the body. As a result, participants are faster to make sensibility judgments when the direction of the response movement matches the direction of the movement implied by the sentence in question. If the sentence is "He opens the drawer", then participants are faster if the 'sensible' (yes) response requires a movement towards the body. This is taken to show that processing the sentences already activates motor regions in the brain; people simulate the action implied in the sentence. And there is interference if one mentally simulates a drawing move, yet has to make a pushing move in response.

In another study participants perform a picture-sentence congruity task (Stanfield, Zwaan and Yaxley 2002). They read the sentence "The eagle is in the sky" or the sentence "The eagle is in its nest", and then view pictures of objects and have to say whether the object depicted was mentioned in the sentence. The picture is either of an eagle with outstretched wings or an eagle with folded wings. It was found that participants respond faster when the content of the picture matches the image the sentence evokes. If the sentence is "The eagle is in the sky" and the picture depicts an eagle with outstretched wings they respond faster as when the picture depicts an eagle with wings folded. This is taken to show that the participants employ perceptual, modality-specific (as opposed to amodal) representations in processing language. These perceptual representations entertain an "analogue relationship" (Stanfield and Zwaan 2001: 153) to what they represent; they represent details of the object that amodal representations would fail to represent, such as, e.g., folded wings.

These studies lend support to the claim that people occasionally engage in simulation when processing language. But this is not quite the claim at issue. The claim is not just that simulation is a byproduct of or occasionally happens to accompany language processing. Rather—as pointed out before—the claim is supposed to be a stronger claim to the effect that simulation is at least a necessary condition for language comprehension.[3]

Some seem to defend an even stronger claim according to which understanding just IS simulation. Vittorio Gallese and George Lakoff, e.g., claim that understanding is imagination; and they further claim that the latter is (mental) simulation (Gallese and Lakoff 2005). Rolf Zwaan and Michael Kaschak have it that in "a very literal sense, the comprehension of a sentence about removing the pie from the oven relies on much the same machinery that would be involved in actually carrying out the action" (Zwaan and Kaschak 2009: 368). In a similar vein, Raymond Gibbs declares that his "personal view is that online language processing is best characterized as a simulation process …" (Gibbs 2005: 87).

---

[3] Of course, the notion of simulation also stands in dire need of clarification (cf., e.g., Sanford 2008).

And with respect to terms expressing emotional states, Glenberg and colleagues, e.g. hold that "understanding of language about emotional states requires that those emotional states be simulated, or partially induced, using the same neural and bodily mechanisms as are recruited during emotional experience." (Glenberg et al. 2005: 120) (The list is not meant to be exhaustive.) But then, given that something like semantic memory (and presumably other faculties as well) is also necessary for language comprehension, one might think that simulation can at best be a necessary condition for understanding—unless it could be shown that semantic memory is located in sensory-motor regions of the brain (on the neural correlates of semantic memory cf., e.g., Binder and Desai 2011).[4]

Moreover, even the weaker claim that simulation is (only) necessary for language comprehension is not uncontroversial. Not only have the results by Hauk, Pulvermüller and others come under attack recently. For example, in a meta-analysis of 29 studies Christie Watson and colleagues conclude that they "did not find evidence for consistent involvement of premotor or motor cortices in the representation of action concepts. At the very least, this finding argues against the hypothesis that premotor or motor cortex activations are inherent to the process of understanding action concepts." (Watson et al. 2013: 1202). What is also missing in order to put the claim that simulation is necessary for comprehension on a firm empirical footing is not so much empirical evidence that there are cases in which there is comprehension as well as simulation but rather evidence that if there is no simulation (sensory-motor activation) then there is no comprehension. Yet that claim is not corroborated by empirical investigation. Rather, there is evidence to the contrary. In discussing various studies with apraxic patients, Bradford Mahon and Alfonso Caramazza reach the conclusion that "cognitive neuropsychological studies of patients with sensory and/or motor impairments demonstrate that such impairments do not necessarily give rise to conceptual deficits." (Mahon and Caramazza 2008: 59). Patients suffering from apraxia are impaired at using objects, but can commonly name the objects (such as a hammer) and say what they are used for. Mahon and Caramazza relate the case of a patient who has been carefully studied and who was "not able to produce any correct pantomime of object use", was "severely impaired at using actual objects correctly" but was "flawless at naming object associated pantomimes…" (Mahonand Caramazza 2005: 483). Production (or simulation) and recognition may not be as closely connected as embodied theories predict.

4. *Abstract terms*

Moreover, so far we have mostly been talking about verbs denoting actions that we can perform. But one might wonder whether we do not also understand verbs denoting experiences or actions that we cannot

[4] Thanks to Markus Kneer for helpful discussion here.

simulate or perform; don't we understand the sentence "The dog's tail began to wag", for example (cf. Hickok 2014)? And, don't we also understand words that do not denote actions or sensory experiences at all? This brings us to the problem of abstract terms, which, in turn, will bring us to the topic of metaphor.

Abstract terms make trouble for embodied approaches to language comprehension because motor or sensory information doesn't seem to be all that relevant to understanding these terms. Yet Laurence Barsalou claims that appearance is deceptive here. He has it that even abstract concepts such as TRUTH "can be represented perceptually" (Barsalou 1999: 600; Barsalou 2009). Jesse Prinz also thinks that all that is needed are modal, perceptual representations. He invites us to

> (c)onsider justice. One way to understand this lofty idea is by grounding it in very concrete scenarios. There are different kinds of injustice, and each can be captured by simulating an event. First, there is inequality. This can be simulated by imagining a situation in which I get two cookies and you get three. Second, there is inequity. For example, you might give me one cookie in exchange for two. Third, there is violation of rights. Suppose I try to eat my cookie and you prevent me from doing so. (Prinz 2012: 129)

Let us gloss over the fact that the situations described are cases of injustice, not justice. Prinz continues:

> We learn the concept by means of very simple cases and then need to figure out whether more complicated cases are sufficiently similar to these (…) Still, the simple scenarios can give us a very concrete idea of what justice is, and that is sufficient for grounding our understanding of this seemingly abstract concept. (Prinz 2012: 129)

Note that Prinz uses "simulation" with a slightly different meaning. "Simulation" here rather means something like "mental imagery", the conscious evoking of a particular scenario. And he also seems to defend a stronger view according to which simulation is not only necessary and also sufficient for language comprehension.

Yet simulating a particular scenario by means of (conscious) mental imagery cannot be a necessary condition for understanding linguistic phrases. Otherwise we would all be very busy vicariously experiencing prior events of injustice while reading the newspaper (there is obviously a cognitive-load-problem lurking in the background). Also, according to Prinz, "[w]e learn the concept by means of very simple cases and then need to figure out whether more complicated cases are sufficiently similar to these" (Prinz 2012: 129). But grasping what various concrete cases have in common is exactly the purpose abstract ideas are supposed to serve; it is their job description. Consequently, abstract ideas come in through the back door. Finally, if all it takes to understand or grasp the concept of justice is to simulate various specific scenarios, then, given that people, arguably, tend to simulate different scenarios, no shared meaning will emerge.

Therefore, I conclude, admittedly without having fully argued the point, that accounts that try to ground abstract ideas "in concrete sce-

narios" fail to explain how comprehension of abstract terms such as "truth" or "justice" can be embodied.[5] Fortunately, there is a very popular account in contemporary cognitive linguistics that is happy to volunteer an alternative explanation.

5. *The Conceptual Theory of Metaphor*

Abstract concepts are metaphorical, the Conceptual Theory of Metaphor, CTM for short, claims. According to George Lakoff, one of the founding fathers of CTM, "everyday abstract concepts like time, states, change, causation, and purpose turn out to be metaphorical". (Lakoff 1993: 203) Mark Johnson, his brother-in-arms, has it that "[a]*ll* theories are based on metaphors because all our abstract concepts are metaphorically defined" (Johnson 2008: 51). But they not only claim that abstract concepts are metaphorical(ly defined). They also claim that metaphor comprehension is embodied!

The idea that abstract terms are metaphorical has intuitive appeal. In using metaphors, we try to understand one kind of phenomenon in terms of another, and we thereby often borrow from the concrete realm of sensory experience in order to conceptualize abstract and mental phenomena (Lakoff and Johnson 1980). We feel blue, complain about her being cold or thin-skinned, we let people "chew over new suggestions and digest new information" (Deutscher 2005: 122), and we say that someone's theory needs more support, is about to collapse, etc. We talk about abstract, 'elusive' things such as emotions, information or theories in terms of more concrete phenomena and mundane activities such as temperature, digestion or buildings.

But CTM is not really a theory about linguistic usage as metaphor is "not a figure of speech but a mode of thought…" (Lakoff 1993: 210). These modes of thought are called conceptual metaphors (hence the name of the theory). They are "mappings across conceptual domains" (Lakoff 1993: 203); more specifically, they are mappings from a source domain which is commonly less abstract onto a target domain which is commonly more abstract. Here is a somewhat overworked example that nonetheless nicely illustrates the point. Let us try to map the domain of journeys onto the domain of love. This is what we get.

---

[5] For a more comprehensive overview over the debate about abstract terms, cf. also, e.g., Dove 2011, 2014; Borghi and Binkofski 2014, or Barsalou and Wiemer-Hasting 2005.

LOVE IS A JOURNEY

Source: journey	Target: love
the travelers	➜ the lovers
the vehicle	➜ the love relationship itself
the journey	➜ events in the relationship
the distance covered	➜ the progress made
the obstacles encountered	➜the difficulties experienced
decisions about the way to go	➜ choices about what to do
the destination of the journey	➜ the goal(s) of the relationship

(Kövecses 2010: 9)

Mappings such as these help us interpret metaphorical utterances. In saying something like "We are at a crossroads", "This isn't going anywhere", or "Look how far we've come" your partner is exploiting the love-is-a-journey mapping. And abstract concepts are metaphorical in that their understanding is also based on conceptual metaphors such as the above one.[6] Yet understanding is achieved not just by mapping elements from one domain onto elements of the other domain. We also 'map' knowledge. We know certain things about journeys, for example, and we use that knowledge in order to understand what love is. We try to understand love in terms of and by what we know about journeys (cf. Lakoff 1993: 206−207). More specifically, we come to employ patterns of inference that we commonly use to reason about journeys in order to reason about love. We know that when one encounters obstacles while travelling or is in danger of getting lost it might be advisable to engage a tour guide. Analogously, one may come to realize that when one encounters difficulties in a love relationship, it might be helpful to engage a psychotherapist as a kind of tour guide, etc. What is mapped are structural relationships. Conceptual metaphors, therefore, come out as analogies,[7] as structure-preserving mappings, an idea that is most clearly expressed in structure-mapping theory as initially developed by Dedre Gentner (1980: 1983).[8]

> The basic intuition is that an analogy is a mapping of knowledge from one domain (the base) into another (the target), which conveys that a system of relations that holds among the base objects also hold among the target

[6] Of course, there are other ways to conceptualize the domain LOVE. Love may be a game, or a plant that needs nurturing, and so on and so forth. Yet Lakoff and Johnson insist that all love-related metaphors are "significantly constitutive of our concept of love" (Lakoff and Johnson 1999: 71–72).

[7] The idea that metaphor is based on analogy goes back at least to Aristotle.

[8] As Dedre Gentner and Brian Bowdle point out, not all metaphors are analogies; some "defy description in terms of alignment", especially certain poetic metaphors such as the following one (if it is a metaphor) from a poem by E. E. Cummings "The voice of your eyes is deeper than all the roses" (cf. Gentner and Bowdle 2008: 110) Also, they defend the 'Career of metaphor hypothesis', according to which "a metaphor undergoes a process of gradual abstraction and conventionalization as it evolves from its first novel use to becoming a conventional 'stock' metaphor." (Gentner and Bowdle 2008: 116).

objects. (Gentner and Clement 1988: 312–313; cf. also Gentner and Bowdle 2008)

Structure is mapped from one domain onto another. And, again, our understanding of abstract terms is said to be based on metaphorical, analogical mappings. Metaphor comprehension, in turn, is said to be embodied. Unfortunately, as soon as we talk about metaphor, conceptions of embodiment proliferate.

6. *Embodied Metaphor*

First, Lakoff and colleagues claim that all complex conceptual metaphors can be decomposed into what they call primary metaphors.[9] Primary metaphors are experientially grounded. An example of a primary metaphor is AFFECTION IS WARMTH.

> For example, for an infant, the subjective experience of affection is typically correlated with the sensory experience of warmth, the warmth of being held. During the period of conflation, associations are automatically built up between two domains. Later, during a period of differentiation, children are able to separate out the domains, but the cross-domain associations persist. (Lakoff and Johnson 1999: 46)

When we were held affectionately as children, we experienced affection and warmth occurring together. We learnt to associate affection with warmth.[10] That helps us to produce and comprehend temperature metaphors ("He greeted me warmly"). According to Lakoff (2012), the association claim is evidenced by various experimental findings. It could be shown, e.g., that subjects holding a cup of hot coffee are prone to evaluate an imaginary individual as warm and friendly—significantly more so than subjects holding a cup of iced coffee (Williams/Bargh 2008). Subjects automatically associate physical warmth with friendliness. This is so, according to Lakoff, "(b)ecause primary metaphors are persistent (long-lasting or permanent) physical circuits in the brain." (Lakoff 2012: 782)

Others claim that metaphor comprehension is embodied in that people understand metaphors via mental imagery. Raymond Gibbs and colleagues, for example, claim that people understand metaphors containing an action verb by imagining themselves engaging in that very action (Gibbs 2006). They conduct psycholinguistic studies in order to corroborate their claim. Mostly, they let participants read metaphorical sentences, the sentence "Let us stomp out racism", e.g., and then

[9] The LOVE IS A JOURNEY metaphor is also built up of various primary metaphors (cf. Lakoff 2008: 26–27). Still, one might wonder whether all complex metaphors are decomposable into primary metaphors and what exactly the principles of (de)composition amount to.

[10] According to Lakoff, thought in general is embodied: "Thought is embodied, that is, the structures used to put together our conceptual system grow out of bodily experience and make sense in terms of it; moreover, the core of our conceptual system is directly grounded in perception, body movement, and experience of a physical and social character" (Lakoff 1987: xiv).

ask them what is "particularly noticeable" (Gibbs and Mattlock 2008: 166) in the image they form. And according to Gibbs, the participants conceive of racism as if it was a physical object–thereby employing the metaphor IDEAS ARE OBJECTS–and then they imagine their bodies in action, imagine themselves stomping (Gibbs and Matlock 2008).

This brings to the fore two new embodiment claims, one according to which

EC: Association is necessary for metaphor comprehension,

and one according to which

EC: Mental imagery is necessary for metaphor comprehension.

But now recall the initial embodiment claim. Applied to the case of metaphor, it comes to this:

EC: Simulation is necessary for metaphor comprehension.

One might think that mental imagery and simulation are simply two sides of the same coin. But although there is, presumably, a close connection here, mental imagery is the conscious evoking of a mental image, while simulation–in the sense at issue here–is (presumably sub-conscious) activation in sensory-motor regions of the brain. Consequently, evidence for a necessary role of simulation in metaphor comprehension ought to come from neuroimaging studies. Yet in a carefully designed fMRI-study by Shirley-Ann Rüschemeyer and colleagues, it could be shown that the comprehension of metaphorical uses of action verbs ("grasp the idea") doesn't yield the same motor activation pattern as the comprehension of literal uses of action verbs ("grasp the cup") (Rueschemeyer et al. 2007). Lisa Aziz-Zadeh and colleagues reached a similar conclusion (Aziz-Zadeh et al. 2006). Others disagree (cf, e.g., Boulenger, Hauk and Pulvermüller 2009). And still others claim that activation in primary motor regions is necessary only for the interpretation of unfamiliar metaphors (Desai et al. 2011). In sum, empirical evidence for an indispensable role of simulation in metaphor comprehension is inconclusive (for a more comprehensive yet still critical assessment, cf., e.g., Casasanti and Gijssels 2015).

But even aside from questions about its empirical support, the claim that metaphor comprehension is embodied either in the association, the mental imagery or the simulation sense, faces a couple of more 'theoretical' problems.

Take association first. It does not seem implausible to assume that association underlies our understanding of temperature metaphors ("She greeted my warmly"). It might also help explain how we understand synaesthetic metaphors such as "The stone statue had a cold smell" (Werning et al. 2006: 2365).[11] Yet note that these metaphors are

---

[11] They define the relevant terms as follows: "A metaphor is synaesthetic if and only if its source domain is perceptual. It is only weakly synaesthetic if its target is not also perceptual and strongly synaesthetic if its target domain, too, is perceptual." (Werning et al. 2006: 2365–2366).

not based on structure-preserving mappings, for there is no structure to be preserved or mapped; they therefore fail to fulfill Lakoff's own characterization. Comprehending metaphors that are based on mappings, on the other hand, requires more than just association. In understanding the LOVE IS A JOURENEY metaphor, e.g., you are not just associating two domains. Rather, you are using knowledge of one domain to better understand the other. Mere association won't do. As a consequence, one might try to distinguish two types of metaphors; those, which are based on association and those which are predominately based on domain knowledge.

Similarly in the case of mental imagery. Imagery might provide a good place to start when it comes to interpreting novel or poetic metaphors. When reading the following two lines

Two roads diverged in the wood, and I
I took the one less traveled by.

of the famous poem "The Road not Taken" by Robert Frost, you might immediately form an image of someone travelling on his own, being alone in the woods and so on. The mental image, by providing you with detailed pictorial and also even affective information about how it feels to be alone in the woods, might help you to grasp the metaphor. Reuven Tsur, one of the founding fathers of cognitive poetics, stresses the role of concrete visual images in the interpretation of metaphor in particular and of poetry in general (Tsur 1999). But what about other metaphors; e.g., the metaphor 'Man is a wolf'? What am I supposed to imagine here? I might imagine a wolf in man's clothing. But in the mind of others the metaphor might conjure up other images, men pack-hunting, for example. Again, domain (encyclopedic) knowledge does most of the work. As Max Black pointed out, the MAN IS A WOLF metaphor "will not convey its intended meaning to a reader sufficiently ignorant about wolves." (Black 1955: 287) Only if you are sufficiently knowledgeable will the wolf-metaphor organize your view of man, as Black puts it (ibid, 288). Again, one might try to distinguish different types of metaphor according to the cognitive processes most likely involved in their interpretation.

Finally, consider simulation again. In order to understand metaphoric expressions such as 'He grasped the idea' or 'man is a wolf', one has to figure out what grasping a cup and grasping an idea, or what man and wolf, have in common. But one also has to understand that an idea is grasped differently than a cup and that man is not a wolf! Otherwise one will not recognize them as metaphors. MAN IS A WOLF is a metaphor exactly because man is NO wolf! And "He grasped the idea" is a metaphorical expression because we grasp an idea in a manner that is somehow similar but also somehow different from the manner in which we grasp a cup! Simulating the action of grasping, thereby activating the same neurons as when one actually grasps a cup, is not what it takes to understand the metaphor. Metaphors draw on simi-

larities–but equally on differences! Understanding metaphors requires that one looks at things differently than one did before! Metaphor, at least novel metaphor, is "a method of *expanding* understanding" (Shapiro 2011: 86). Simulation, if at all, provides us with a theory of literal interpretation. But interpreting metaphors requires that we leave the literal meaning behind.

Now recall that CTM not only claims that metaphor comprehension is embodied but also that conceptual metaphors help us understand, are even "significantly constitutive" of, abstract concepts. Yet that claim, as has often been pointed out, is also fraught with problems. Let me mention just one (for some more, cf., e.g., Kompa forthcoming). CTM makes wrong predictions about the way in which abstract concepts are best learned. "What is love?" the little boy asks. "A journey", his mother replies. Yet the boy is not meant to believe that love *is* a journey. Rather, he is supposed to believe (and has to learn) other things about love; that if two people are in love they like each other a lot, and want to spend time together, and so on. The 'defining' properties need to be learned independently of the mapping, it seems. And only distinct domains can be mapped onto each other.[12] As Barsalou puts it: "Although metaphor most certainly plays a major role in elaborating and construing abstract concepts, it is not sufficient for representing them…" (Barsalou 1999: 600; cf. also, e.g., Murphy 1996)

7. *Summary*

To sum up, abstract terms are not metaphorical, nor is metaphor comprehension embodied. Or at least, neither association nor mental imagery nor simulation seem to play an indispensible role in metaphor comprehension. Still, embodied approaches are right in claiming **a** role for simulation, mental imagery or association in language comprehension in general and metaphor comprehension in particular. It seems highly plausible that on hearing a particular utterance a listener will occasionally simulate a previous experience or actively imagine a scenario that somehow fits the utterance. This might provide her with more detailed, pictorial information about a possible way the world might be if the utterance is correct. Similarly, in the case of a metaphorical utterance. A listener will occasionally conjure up certain images that might help her explore the metaphorical mapping further, especially in the case of novel metaphors.[13] In other cases, association might provide a

[12] Domain knowledge usually precedes metaphorical mapping. As Ellen Winner and Howard Gardner, e.g., point out, a child's ability to interpret metaphor is constrained only by what they know about the domains in question: "That is, there are not inherent limits on the kinds of similarities children can perceive. All that is necessary is sufficient knowledge of the domains involved" (Winner and Gardner 1993: 427).

[13] Evidence concerning the neural correlates underlying the processing of novel metaphors is somewhat inconclusive, especially with respect to the role of the right

route to comprehension. But there is no reason to think that every time a linguistic item, let alone a metaphor, is being processed simulation, association, or imagery has to take place. Still, as suggested above, one might try to distinguish different types of metaphors according to the cognitive processes most likely involved in their comprehension, even if this will hardly yield a clear-cut distinction. So, in sum, I side with those who claim that in language comprehension in general, and metaphor comprehension in particular sensory-motor (as well as affective) information is appealed to in a context-sensitive, task-dependent manner (cf., e.g., Desai et al 2011,[14] Desai et al. 2013;[15] or van Dam et al. 2014?[16]).

References

Adams, F. 2010. "Embodied Cognition." *Phenomenology and the Cognitive Sciences* 9: 619–628. doi:10.1007/s11097–010–9175–x.

Aziz-Zadeh, L. et al. 2006. "Congruent Embodied Representations for Visually Presented Actions and Linguistic Phrases Describing Actions." *Current Biology* 16: 1818–1823. doi:10.1016/j.cub.2006.07.060.

Barsalou, L.W. 1999. "Perceptual symbol systems." *Behavioral and Brain Sciences* 22: 577–609.

Barsalou, L. 2009. "Situating Concepts." In P. Robbins and M. Aydede (eds.) 2009: 236–263.

Barsalou, L.W. and Wiemer-Hastings, K. 2005. "Situating abstract concepts". In D. Pecher and R. Zwaan (eds.). 2005: 129–163.

Binder, J. R. and Desai, R. H. 2011. "The Neurobiology of semantic Memory." *Trends in Cognitive Science* 15 (11): 527–536. doi:10.1016/j.tics.2011.10.001.

Black, M. 1955. "Metaphor." *Proceedings of the Aristotelian Society* 55: 273−294.

hemisphere (yet cf., e.g., Mashal et al. 2007).

[14] Desai et al. (in an fMRI-study) investigated participants' responses to literal action sentences, metaphoric sentences, and abstract sentences of varying familiarity. They "hypothesized that relatively unfamiliar (literal and metaphoric) action language engages the sensory-motor systems because comprehension of such expressions involves relatively detailed simulations of literal actions." (Desai et al. 2011: 2377). Their results confirmed the hypothesis.

[15] According to Rutvik Desai and colleagues, one "question is whether the involvement of sensory-motor information is obligatory (because it is an essential part of semantic representation) or context-dependent (varying with factors such as task-demands or expectations due to the nature of the stimuli)." (Desai et al 2013: 1–2) As they point out, their results suggest that the involvement of the sensory-motor system decreases as abstraction increases, "highlighting the context-sensitive nature of semantic processing" (Desai et al 2013: 1).

[16] Wessel van Dam and colleagues investigated at what level of language processing sensorimotor activity comes into play. They reach the conclusion that "… recent studies have shown that sensorimotor information is recruited in a flexible manner during language comprehension…" (van Dam et al. 2014: 407).

Borghi, A. M. and Binkofski, F. 2014. *Words as social Tools: An Embodied View on Abstract Concepts*. New York: Springer.

Boulenger, V., Hauk, O. and Pulvermüller, F. 2009. "Grasping Ideas with the Motor System: Semantic Somatotopy in Idiom Comprehension." *Cerebral Cortex* 19 (8): 1905–1914.

Casasanto, D. and Gijssels, T. 2015. "What makes a metaphor an embodied metaphor?" *Linguistic Vanguard*. doi:10.1515/lingvan-2014–1015.

Deutscher, G. 2005. *The Unfolding of Language*. London: Arrow Books.

Desai, R. H. et al. 2011. "The neural career of sensory-motor metaphors." *Journal of Cognitive Neuroscience* 23 (9): 2376–2386.

Desai, R. H. et al. 2013. "A piece of the action: Modulation of sensory-motor regions by idioms and metaphors." *Neuroimage* 83. doi:10.1016/j.neuroimage.2013.07.044 (NIH Public Access; Author Manuscript).

Dove. G. 2011. "On the need for embodied and dis-embodied cognition." *Frontiers in Psychology* 1: 1–13. doi:10.3389/fpsyg.2010.00242.

Dove, G. 2014. "Thinking in Words: Language as an Embodied Medium of Thought." *Topics in Cognitive Science* 6: 371–389. doi:10.1111/tops.12102.

Feldmann, J. and Narayanan, S. 2004. "Embodied meaning in a neural theory of language." *Brain and Language* 89: 385–392.

Gallese, V. and Lakoff, G. 2005. "The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge." *Cognitive Neuropsychology* 22 (3): 455–479. doi:10.1080/02643290442000310.

Gentner, D. 1980. *The structure of analogical models in science (Tech. Rep. No. 4451)*. Cambridge: Bolt Beranek & Newman.

Gentner, D. 1983. "Structure-mapping: A theoretical framework for analogy." *Cognitive Science* 7: 155–170.

Gentner, D. and Bowdle, B. F. 2008. "Metaphor as Structure Mapping." In R. W. Jr. Gibbs (ed.) 2008: 109–128.

Gentner, D. and Clement, C. 1988. "Evidence for relational selectivity in the interpretation of analogy and metaphor." In G. H. Bower (ed.) *The psychology of learning and motivation: Advances in research and theory*, vol. 22. New York: Academic Press: 307–358.

Gibbs, R. W. 2005. "Embodiment in Metaphorical Imagination." In D. Pecher and R. Zwaan, (eds.) 2005: 65–92.

Gibbs, R. W. Jr. 2006. "Metaphor interpretation as embodied simulation." *Mind & Language* 21: 434–458.

Gibbs, R.W. Jr. (ed.). 2008. *The Cambridge Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press.

Gibbs, R. W. Jr. and Matlock, T. 2008. "Metaphor, imagination, and simulation: Psycholinguistic evidence." In R.W. Jr. Gibbs (ed.) 2008: 161–176.

Glenberg, A. and Kaschak, M. 2002. "Grounding language in action." *Psychonomic Bulletin and Review* 9: 558–565.

Glenberg, A. et al. 2005. "Grounding Language in Bodily States. The Case for Emotions." In D. Pecher and R. A. Zwaan (eds.) 2005: 115–128.

Hauk, O., Johnsrude, I. and Pulvermüller, F. 2004. "Somatotopic Representation of Action Words in Human Motor and Premotor Cortex." *Neuron* 41: 301–30.

Hickok G. 2014. *They Myth of Mirror Neurons*. Ney York: W.W. Norton and Company Inc.

Johnson, M. 2008. "Philosophy's dept to metaphor." In R. W. Jr. Gibbs (ed.) 2008: 39–52.

Kaschak, M. P. et al. 2014. "Embodiment and Language Comprehension." In L. Shapiro (ed.) 2014: 118–126.

Kompa, N. Forthcoming. "Insight by Metaphor: the Epistemic Role of Metaphor in Science." Forthcoming in A. Heydenreich and K. Mecke (eds.). *Physics & Literature: Theory – Popluarization – Aestheticization*. Berlin: De Gruyter.

Kövecses, Z. 2010. *Metaphor: A Practical Introduction*. Oxford: Oxford University Press.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories reveal about the Mind*. Chicago: The University of Chicago Press.

Lakoff, G. 1993. "The contemporary theory of metaphor." In A. Ortony (ed.). *Metaphor and Thought*. Cambridge: Cambridge University Press: 202–251.

Lakoff, G. 2008. "The neural theory of metaphor." In R.W. Jr. Gibbs (ed.) 2008: 17–38.

Lakoff, G. 2012. "Explaining Embodied Cognition Results." *Topics in Cognitive Science* 4: 773–385.

Lakoff, G. and Johnson, M. 1980. *Metaphors we live by*. Chicago: The University of Chicago Press.

Lakoff, G. Johnson, M. 1999. *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.

Mahon, B. Z. and Caramazza, A. 2005. "The Orchestration of the Sensory-Motor System – Clues from Neuropsychology." *Cognitive Neuropsychology* 22 (3–4): 480–494.

Mahon, B. Z. and Caramazza, A. 2008. "A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content." *Journal of Physiology* 102: 59–70.

Mashal, N. et al. 2007. "An fMRI investigation of the neural correlates underlying the processing of novel metaphorical expressions." *Brain and Language* 100: 115–126.

Murphy, G. L. 1996. "On metaphoric representation." *Cognition* 60: 173–204.

Pecher, D. and Zwaan, R. A. (eds.). 2005. *Grounding Cognition: The Role of Perception and Action in Memory, Language and Thinking*. Cambridge: Cambridge University Press.

Prinz, J. 2012. *Beyond Human Nature*. London: Penguin Books.

Pulvermüller, F. 2005. "Brain mechanisms linking language and action." *Nature Reviews Neuroscience* 6: 576–582.

Robbins, P. and Aydede, M. 2009b. "A Short Primer on Situated Cognition." In P. Robbins and M. Aydede (eds.) 2009: 3–10.

Robbins, P. and Aydede, M. (eds.) 2009. *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press.

Rüschemeyer, S.-A., Brass, M. and Friederici, A. D. 2007. "Comprehending Prehending: Neural Correlates of Processing Verbs with Motor Stems." *Journal of Cognitive Neuroscience* 19 (5): 855–865.

Sanford, A J. 2008. "Defining Embodiment in understanding." In M. de Vega, A. Glenberg, and A. C. Graesser (eds.). *Symbols and Embodiment.*

Debates on Meaning and Cognition. Oxford: Oxford University Press: 181–194.

Scorolli, C. 2014. "Embodiment and Language." In L. Shapiro (ed.) 2014: 127–138.

Shapiro, L. 2011. *Embodied cognition.* New York: Routledge.

Shapiro, L. (ed.) 2014. *The Routledge Handbook of Embodied Cognition.* London: Routledge.

Stanfield, R. A. and Zwaan, R. A. 2001. "The Effect of Implied Orientation Derived from Verbal Context on Picture Recognition." *Psychological Science* l12 (2): 153–156

Stanfield, R. A., Zwaan, R. A. and Yaxley, R. H. 2002. "Language Comprehenders mentally represent shapes of Objects." *Psychological Science* 13 (2): 168–171.

Tsur, R. 1999. "Lakoff's Road not Taken." *Pragmatics & Cognition* 7 (2): 339–359.

van Dam, W. O. et al. 2014. "Flexibility in Embodied Language Processing: Context Effects in Lexical Access." *Topics in Cognitive Science* 6: 407–424. doi:10.1111/tops.12100.

Watson, C. E. et al. 2013. "Action Concepts in the Brain: An Activation Likelihood Estimation Meta-analysis." *Journal of Cognitive Neuroscience* 25 (8): 1191–1205.

Winner, E. and Gardner, H. 1993. "Metaphor and Irony: Two Levels of Understanding." In Ortony, A. (ed.). 1993 [1979]. *Metaphor and Thought.* Cambridge: Cambridge University Press: 425–443.

Werning, M., Fleischhauer, J. and Beseoglu, H. 2006. "The cognitive accessibility of synaesthetic metaphors." In Sun, R. and Miyake, N. (eds.). *Proceedings of the Twenty-eight Annual Conference of the Cognitive Science Society.* London: Lawrence Erlbaum Associates: 2365–2370.

Williams, L. E. and Bargh J. 2008. "Experiencing Physical Warmth Promotes Interpersonal Warmth." *Science* 322 (5901): 606–607. doi:10.1126/science.1162548.

Zwaan, R. A. and Kaschak, M. P. 2009. "Language in the Brain, Body, and World." In Robbins, P. and Aydede, M. (eds.) 2009: 368–381.

Baseless Knowledge

GUIDO MELCHIOR
Department of Philosophy, University of Graz, Graz, Austria

It is a commonly held view in contemporary epistemology that for having knowledge it is necessary to have an appropriately based belief, although numerous different views exist about when a belief's base is appropriate. Broadly speaking, they all share the view that one can only have knowledge if the belief's base is in some sense truth-related or tracking the truth. Baseless knowledge can then be defined as knowledge where the belief is acquired and sustained in a way that does not track the truth. I will argue that rejecting baseless knowledge leads to controversial consequences. The problem increases if we consider contrasting persons who know because of appropriate belief forming processes but who fail to possess further epistemic virtues such as understanding. I will not argue which belief bases constitute a sufficient condition for knowledge. Rather I will stress the point that the common assumption that an appropriate basing relation constitutes a necessary condition for knowledge has controversial consequences.

1. *Baseless knowledge*

Overview: First, I will present two examples of persons to whom it is *prima facie* not implausible to ascribe knowledge, although their beliefs are not appropriately based. Second, I will show that the claim that these persons do not know becomes even more controversial if we consider contrasting persons to whom knowledge accounts typically ascribe knowledge that assume a basing relation as necessary for knowing. Third, I will illustrate in more detail, why prominent externalist knowledge accounts are committed to strongly affirming that baseless knowledge does not exist. Those accounts that I will discuss are process-reliabilism, sensitivity, safety and virtue epistemology. I will argue that these accounts not only face a problem if baseless knowledge clearly exists, they have already counter-intuitive consequences if the

existence of baseless knowledge is *prima facie* disputable. Fourth, I will compare my cases to Lehrer's gypsy lawyer case and, fifth, I will argue why we cannot understand the presented cases of baseless knowledge as instances of causal overdetermination or pseudo-overdetermination as one might suggest.

Here are two cases where it seems controversial to deny knowledge although the beliefs are not appropriately based:

Case 1: The obsessed detective

> Inspector X is an ambitious and passionate detective at a police department. Mrs Charming has been murdered and X is commissioned to catch her murderer. X visits her home to meet her husband Mr Charming. At the very moment of seeing Mr Charming for the first time X has the intuition that Charming murdered his wife. X becomes immediately convinced that Charming is the murderer, although Charming is a very handsome, distinguished and popular person and nobody else but X believes that he could have committed the crime. X starts seeking for evidence that Charming murdered his wife, but since Charming is also smart, X cannot find any piece of evidence for years. Over the years, X becomes obsessed with this case. She is totally convinced that Charming killed his wife and nothing, not even evidence to the contrary, could change her conviction anymore. After years, X finds the gun that is evidently the murder weapon and there are only Charming's fingerprints on it. X has hereby proven that Charming murdered his wife. She can easily convince all her colleagues and the rest of the world. X gets a promotion and becomes a legend of her department.

Case 2: The obsessed scientist

> Since she was a young boy O wanted to become a scientist like her hero Albert Einstein. O fulfils her dream and becomes professor of physics. However, O is not as successful as she dreamt of as a young boy. Like many colleagues, she is investigating a physical phenomenon ϕ that has not been explained yet, but without any success. One evening, O is kissed by the muse and has the intuition that there exists an undiscovered subatomic particle, whose features explain ϕ. O is enthusiastic about this idea, since the particle could be named after her, she would become immortal in the scientific community, and her childhood dream would come true. However, the particle is hard to find and O gets more and more obsessed with proving its existence until her obsession reaches a point, where nothing, not even evidence to the contrary, could change her conviction anymore. Finally, O can perform the decisive experiment that proves the existence of the sought-after particle. The particle becomes called the O-particle and O receives the Nobel prize for this discovery.[1]

---

[1] Klein (2012) presents a similar example of an astronomer who acquires a true belief about the moon by misunderstanding a conversation between two students, but then successfully proves that the believed proposition is true.

Further cases of obsessed believers can easily be created. Inspector X has proven that Charming murdered his wife and she received all the merits for her achievement. Professor O has proven the existence of the sub-atomic particles and was awarded the Nobel-prize. Does X *know* that Charming murdered his wife? Does O *know* that the particles exist? I think one intuition is that they do not know before finding evidence, but that they know after having found evidence. Accordingly, adequate theories of knowledge should capture this intuition. Remarkably, nearly all contemporary accounts on knowledge are committed to claiming that it is clearly the case that they still do not know after having found evidence.

I will call any kind of reason or cause for holding or sustaining a belief its *base*, including mental states as well as external factors such as states, events or processes that cause a belief.[2] I defined *baseless knowledge* as knowledge where the belief is acquired and sustained in a way that does not track the truth. According to this definition, knowledge of obsessed believers as in the case of inspector X and professor O is an instance baseless knowledge.[3]

In what follows, I will mainly discuss the case of professor O. Obviously, we obtain the same results for detective X. O has a strong conviction and possesses clear evidence. However, her beliefs are only based on her intuitions and not on this evidence. One reason why we seek evidence or justification is that we want to base new beliefs on it. In this respect, persons who are already totally convinced about a proposition, but still keep seeking for evidence for it might be practically incoherent. However, there are epistemic contexts, as in the two examples of the inspector and the scientist, where we seek evidence for the purpose of *proving* the truth of our convictions *to others*, but not for basing our own beliefs on it. In these contexts, which involve a social component, it is rational to seek evidence, even after having acquired an unchangeable conviction.

Let's suppose further that O has strong intuitions and acts according to them but that her intuitions are not truth-conducive in any sense, i.e. her intuitions are not more likely to be true than mere guessing.[4] Under this assumption, the following holds:

R: O's intuitions are not reliable belief forming processes.

R can be strengthened by assuming that O's intuitions are even *unreliable* belief forming processes such that mere guessing is more likely to be true than her intuitions.

Moreover, O is a totally obsessed believer and would not change her beliefs under any circumstances. This obsession can be characterised by the following counterfactual claims:

---

[2] Here, I follow the terminology of Williamson (2000, and 2009).

[3] The notion of baseless knowledge is used by Turri (2011).

[4] I do not assume a particular concept of intuitions here.

CF_1:   O would believe that the particles exist, if they did not exist.
CF_2:   O would believe that the particles exist, if she did not perform any verifying experiment.
CF_3:   O would believe that the particles exist, if she or someone else did perform an experiment that proves the contrary.

The counterfactual conditional CF_1 is true, since O holds her belief only because of her intuitions. Her belief is not caused by the facts that make her belief true. CF_2 is true because the intuitions are completely causally independent from the process of seeking and finding evidence.

CF_3 is normally a stronger claim than CF_2 because persons who believe that p for a reason despite evidence against it usually also believe that p, if there is no evidence. However, the opposite implication does not hold. CF_3 holds, if O is so obsessed with her convictions that she is psychologically unable to revise her views. This can be the case if proving the truth of her intuitions is for some psychological reason so important to her that she would go insane if she did find evidence that her convictions are false.[5]

O is prejudiced. She would hold her belief without any evidence and she would even sustain her belief despite evidence to the contrary. In this respect, she violates two rules of rationality: first, that one should believe only if one has reasons or justification for believing, and second, that one should not believe if one has evidence to the contrary. Violating the second rule is more serious than violating the first one. In this respect, O is not an ideal rational agent. However, in evaluating the epistemic circumstances of O it is important to note that CF_1-CF_3 are only *counterfactual* conditionals. Whenever we consider them, we consider something that is not the case. In the *actual* world, O proved that the subatomic particles exist. In the *actual* world, she does not believe despite her beliefs being false, her finding no evidence or her finding evidence to the contrary. Considering different possible worlds for evaluating the actual world, is, of course, a usual move in philosophy. However, I doubt that we would do this outside philosophical contexts in clear cases of praiseworthy discoveries. It seems implausible that in non-philosophical contexts we determine whether Albert Einstein knows that $E = mc^2$ by asking whether he believes this in another possible world where he did not find the proof or did find a proof to the contrary. I think in non-philosophical contexts we do not put that much emphasize on counterfactual situations.

The counterfactual claims CF_1-CF_3 aim at illustrating that there is no appropriate causal relation between the beliefs and their truthmakers or between the evidence and the belief. However, one could object that we cannot adequately capture causal relations by using counterfactual claims. Taking this objection into account, we can reformulate

---

[5] Moreover, I assume in the presented cases that the beliefs that the detective and the scientist form are not based on any kind of inference to the best explanation that might involve a truth connection.

the case of O such that she simply does not have good enough reasons to believe, based on an intuitive understanding of good enough reasons.

O bases her belief on a defective belief forming process. However, do we really want to admit that she does not know? If O knows, her knowledge is not ideal with respect to the involved belief forming processes. If one accepts *fallible* justification or warrant, then warrant and justification come in degrees. We have infallible justification on one end of the spectrum and the weakest possible evidence that still converts true beliefs into knowledge on the other end, and numerous forms of justification or warrant in between. In this case, we are inclined to accept non-ideal knowledge with respect to justification and warrant. But if we accept non-ideal knowledge in one respect, why should we exclude non-ideal knowledge in another respect, i.e. with respect to the belief forming process? In what follows, I will present various counterintuitive claims that follow from the assumption that O does not know.

Those who deny that O knows argue that she fails to know because of the defectiveness of her belief forming processes. If one accepts this view, then counterintuitive general claims of the following form can be true:

Ci_1: S is convinced that p and S has proven that p, but S does not know that p.

In case of O, the following is true:

Ci_{1O}: O is convinced that O-particles exist and O has proven it, but O does not know it.

In this case, one has to accept that O believes a proposition and has proven the proposition to be true, *and* she does not know this proposition. One can increase the counter-intuitiveness of Ci_1 and of the following claims by adding that S has not only proven that p, but that S is also aware of this fact.

Moreover, the counter-intuitiveness increases, if we consider the fact that it is O's *achievement* or *merit* that she has proven her beliefs to be true (or that O is creditable for having proven that O). O did not stumble luckily across the evidence, but was searching intentionally and systematically for it. If she does not possess knowledge, then counterintuitive claims of the following form can be true as well:

Ci_2: S is convinced that p and S has proven that p and it is S's merit that S has proven that p, but S does not know that p.

In the case of O, the following is true:

Ci_{2O}: O is convinced that O-particles exist and O has proven it, and this is her merit, but O does not know it.

Claims of type Ci_2 sound even more implausible than those of type Ci_1.

2. *Contrasting persons*

The claim the neither the inspector nor the scientist knows becomes even more problematic, if we contrast them with other persons to whom

those accounts ascribe knowledge, which at the same time deny that the inspector or the scientist know due to inappropriate belief forming processes. In the following, I will first contrast them with other experts and, second, with laypersons.

Let's suppose that O has a colleague, professor P. O performs her experiment and convinces P that the sub-atomic particles exist and that they explain the physical phenomenon ϕ. As a result, P believes that O-particles exist. P is justified in believing it and P believes it because of the evidence. Therefore, those who argue that O does not know because her belief is not based on the evidence presumably admit that P knows.[6] If one agrees with this view, then claims of the even more counterintuitive type can turn out to be true:

Ci_3: S_1 is convinced that p and S_1 has proven that p to S_2, but S_1 does not know that p, whereas S_2 does.

In the case of O, the following is then true:

Ci_{3O}:   O is convinced that O-particles exist and O has proven this to professor P, but O does not know it, whereas P does.

In these cases, the persons who present the proof do not know, but their colleagues, to whom the evidence is presented, know. This seems controversial.

The cases of contrasting persons can be strengthened by taking *laypersons* as contrasting persons into account. Take for example schoolboy B who learns from a textbook that O-particles exist.[7] If we admit that schoolboy B acquires knowledge this way because his beliefs are appropriately based, then the following counter-intuitive claims are also true:

Ci_{3O}:   O is convinced that O-particles exist and O has proven it. Schoolboy B learns from a textbook that O-particles exist, but O does not know it, whereas B does.

These claims sound even more implausible. We can increase their implausibility by assuming that O has the strongest possible evidence and understanding and the contrast person the weakest possible justification and understanding that turns an appropriately formed true belief into knowledge.

Moreover, we can suppose that B and everybody else would not know that O-particles exist, if O hadn't proven it because if O hadn't proven it nobody else would have proven it. In this respect, we can say that it is O's merit that B knows that O-particles exist. In this case, instances of the following general claim are true:

---

[6] We can construct a similar scenario for X, when she convinces colleague Y via presenting the evidence.

[7] Analogously we can assume Granny G as contrasting layperson for inspector X, who is reading in the local newspaper that X has proven that Charming murdered his wife and believes this because of reading the newspaper.

Ci_4: S_1 is convinced that p and S_1 has proven that p to S_2, and S_2 thereby comes to know that p and it is S_1's merit that S_2 knows that p, but S_1 does not know that p.

In the case of O, the following is true:

Ci_{4O}:   O is convinced that O-particles exist and O has proven it and it is O's merit that schoolboy B knows that O-particles exist, but O does not know it.[8]

These claims seem very controversial. To make this point more explicit, consider the following dialogue between two physicists at the Nobel Prize ceremony:

A: O really deserves the Nobel Prize. Nobody believed that the existence of a particle can explain the ϕ-phenomenon. She was seeking for decisive evidence for years. Thanks to her, we now understand this phenomenon and know that O-particles exist.

B: That's true; it's just a personal tragedy that O will never know all this.

A: Why not?

B: Because she became obsessed with the idea that these particles exist during her research.

A: I see. What a pity.

We do not easily admit that those two physicists share a correct understanding of knowledge. Rather we are at least inclined to affirm that this is not clearly the case.[9]

3. *Externalist knowledge accounts*

Externalist accounts of justification and knowledge share the view that appropriately forming or sustaining a belief is at least necessary for

---

[8] Lackey (2007 and 2009) argues that S can know that p without deserving credit for truly believing that p. If one goes with Lackey, then one can also suppose that schoolboy B knows without even deserving credit for it, which is a further contrast to the scientist, who deserve credit that B knows. Moreover, Lackey argues that a person S can know that p via testimony from a person T, although T does not know that p. Lackey (2008) presents the case of Stella, a teacher who believes in the truth of creationism and in the falsity of evolutionary theory. Nevertheless, Stella carries out her duty and teaches evolutionary theory at school. Lackey argues that Stella does not know that evolutionary theory is true, although her students can know via Stella's reliable testimony that evolutionary theory is true. Thus, one might think that the case of professor O and schoolboy B is just an instance of Lackey's case and not particularly problematic. Notably, the two cases are different. Stella fails to know that evolutionary theory is true because she fails to *believe* it. O, in contrast, believes that O particles exist and acknowledges the evidence for it. O just fails to believe for the 'right' reason.

[9] Notably A and B take O's strange belief-forming process explicitly into account. Still, their rejection of knowledge seems false. Hence, the counter-intuitiveness of Ci1O-Ci4O is not based on the fact that they do not mention O's strange belief-forming process.

having knowledge. Process reliabilism is the view that beliefs have to result from processes that reliably produce true beliefs, sensitivity and safety principles interpret the belief forming process modally, and virtue epistemologists claim that the belief has to result from an agent's epistemic virtues. In what follows, I will argue briefly, why externalist accounts are committed to strongly affirming Ci_1 and Ci_2 and presumably also Ci_3 and Ci_4. I will not offer a complete selection of all externalist accounts defending a basing relation as a necessary component of knowledge; I only focus on some of the most prominent, selecting the best-known representatives of these accounts.

3.1. *Knowledge without reliability*

Process reliabilism is (or at least was) more concerned with epistemic justification than with knowledge. Goldman (1979: 9) argues that "correct principles of justified beliefs must be principles that make causal requirements, where "cause" is constructed broadly to include sustainers as well as initiators of beliefs (i.e., processes that determine, or help to overdetermine, a belief's continuing to be held)." Goldman notes that belief-forming processes that are intuitively justification-conferring share reliability, but that faulty belief forming processes do not. Goldman (1979: 10) concludes that the "justificational status of a belief is a function of the reliability of the process or processes that cause it, where [...] reliability consists in the tendency of a process to produce beliefs that are true rather than false." The first version of reliability that Goldman considers is the following:

> S's believing p at t is justified if and only if S's believing p at t results from a reliable belief-forming process (or set of processes).

This version captures the core idea of process reliabilism. O holds and sustains her beliefs as a result of her intuitions, which are not reliable belief forming processes. Therefore, these beliefs are not justified according to process reliabilism. Hence, Goldman's account implies the counter-intuitive claims Ci_1 and Ci_2 reformulated for justified beliefs. Moreover, suppose that the processes leading to schoolboy B's belief are reliable since the textbooks are reliable sources. Hence, reliabilists would admit that B has justified beliefs and that the counterintuitive claims Ci_3 and Ci_4 reformulated for justified beliefs hold as well.

Goldman (1979) notes that beliefs may be over-determined in the sense that they may have a number of distinct ancestral trees. He admits that they need not all be full of reliable or conditionally reliable processes, but at least one ancestral tree must have reliable or conditionally reliable processes throughout. However, the case of professor O can easily be formulated in way that her belief is not causally over-determined, since it is exclusively caused and causally sustained by her intuitions, which are not reliable processes. Hence, there is no ancestral tree that has reliable processes throughout.

The only possible way for reliabilism to avoid the conclusion that O does not have justified beliefs that comes to my mind is to claim that O acquires a *second* belief when performing the experiment for the first time. This second belief would be the result of reliable belief-forming processes, and, hence they would be justified. But this is not a viable way for reliabilists: If they claim ad hoc that anybody with appropriate evidence acquires a second belief, then these beliefs are not primarily characterized by the quality of the belief forming processes anymore, but by the quality of the evidence the person has.

Goldman regards his original version of reliabilism as unsatisfactory and later refined it in various ways by adding further conditions as necessary for justified believing.[10] However, the belief forming processes of O are simply unreliable and, therefore, they violate the core idea of process reliabilism. They fail to fulfil at least one necessary condition for justified believing, and in some cases, they fail to fulfil additional conditions as well. Hence, these refined versions of reliabilism also imply the truth of the counter-intuitive claims Ci_1 and Ci_2 for justified beliefs. Moreover, one can easily modify the case of B in a way that B's belief fulfils all the conditions for justified believing. As a result, these refined versions of process reliabilism also imply Ci_3 and Ci_4 for justified beliefs.[11]

3.2. *Knowledge without sensitivity*

Nozick (1981) interprets knowledge modally. As a first approximation, he defines that S knows p iff (1) p is true; (2) S believes that p; (3) If p weren't true, S wouldn't believe that p; and (4) If p were true, S would believe that p. Each of the premises (1)–(4) constitute a necessary condition for knowledge. Condition (3), which is the crucial one for most purposes, is often called the sensitivity requirement. In a further step, Nozick (1981: 179) defines knowing via a method as follows:

S knows, via method (or way of believing) M, that p:

---

[10] Goldman (1979: 20) adds as a further condition for justification to mere reliability that "there is no reliable or conditionally reliable process available to S which, had it been used by S in addition to the process actually used, would have resulted in S's not believing p at t." In *Epistemology and Cognition* (1986: 111f), Goldman adds a negative higher-order condition in the form of a non-undermining condition according to which a cognizer cannot be justified, if she does have reason to believe that her first-order belief *isn't* reliably caused. For further variants of process reliabilism and its connection to related accounts see Goldman (2011).

[11] Plantinga (1993) regards reliability as a necessary, but not as a sufficient condition for justification or for the third condition of knowledge beyond true belief, which he calls "warrant". He introduces the notion of proper function which implies the existence of a design plan. Bergmann (2006) offers a modified version of Plantinga's account of warrant, but also defends reliability as necessary for knowledge. Hence, those following Plantinga or Bergmann must accept Ci1 and Ci2 as well.

(1) p is true.
(2) S believes, via method or way of coming to believe M, that p.
(3) If p weren't true and S were to use M to arrive at a belief whether (or not) p, then S wouldn't believe, via M, that p.
(4) If p were true and S were to use M to arrive at a belief whether (or not) p, then S would believe, via M, that p.

Notably, Nozick explicitly deals with the question of causal over-determination, which is relevant for the case of O. He discusses the case of a father, who believes that his son is innocent of committing a particular crime via two methods, namely because of faith in his son and because of a conclusive demonstration in the courtroom.[12] This case is similar to the ones of obsessed knowledge. Nozick argues that the father does only know if the good method outweighs the defective method. If persons come to believe that p via two or more distinct methods, then the modal features of the "dominant" method that outweighes all other methods determines whether the person knows. Nozick (1981: 182) captures this idea as following:

> S knows that p if and only if there is a method M such that (a) he knows that p via M, his belief via M that p satisfies conditions 1–4, and (b) all other methods M_1 via which he believes that p that do not satisfy conditions 1–4 are outweighed by M.

Nozick (1981: 182) defines outweighing methods as following:

> [M]ethod M is outweighed by others if when M would have the person believe p, the person believes not-p if the other methods would lead to the belief that not-p, or when M would have the person believe not-p, the person believes p if the other methods would lead to the belief that p.

The case of the scientist O is set up in way that there is no causal relation between O's belief and the evidence for this belief. However, let's suppose, for the sake of the argument that O believes that the particles exist via her intuition *and* via the experiment. Her belief via the experiment satisfies Nozick's conditions 1–4.[13] Hence, O knows *via the experiment*. O knows according to Nozick, if and only if all other methods via which she believes that the particles exist are outweighed by the experiment. O also believes via her intuition, but her intuition clearly violates conditions (3) and (4): If the particles did not exist and O were to use her intuition to arrive at a belief whether (or not) they exist, then S would still believe, via her intuition, that they exist. Therefore, condition (3) is not satisfied. Therefore, O only knows, according to Nozick, if

---

[12] Nozick refers for this example to Armstrong (1973).

[13] This is the case because (1), the theory is true, (2), O believes the theory via the experiment, (3), if the theory weren't true and O were to use the experiment to arrive at a belief whether (or not) the theory is true, then O wouldn't believe, via the experiment, that the theory is true, and, (4), if the theory were true and O were to use the experiment to arrive at a belief whether (or not) the theory is true, then S would believe, via the experiment, that the theory is true.

her intuition is outweighed by the experiment. This would be the case if: when O's intuition would lead to the belief that the particles exist, O would nevertheless not believe this if the experiment showed that such particles do not exist. However, this condition is not satisfied for O since she would believe according to what her intuitions tell her, no matter what results the experiment gains. Therefore, O's intuitions are not outweighed by the experiment and O does not know according to Nozick. Thus, Nozick's definition of knowledge implies the truth of Ci_1 and Ci_2.

Moreover, one can easily reformulate the case of schoolboy B in a way that his beliefs fulfil Nozick's refined concept of knowledge, either by supposing that he only believes via one method which is reading the textbook or by supposing that all other methods are outweighed by this method. In these cases, the counterfactual conditionals Ci_3 and Ci_4 hold as well.

3.3. *Knowledge without safety*

Sensitivity accounts of knowledge have been criticised in various ways.[14] Sosa (1999) suggests replacing sensitivity by the alternative modal principle *safety*, which he defines as following:

> Call a belief by S that p "safe" iff: S would believe that p only if it were so that p.
> (Alternatively, a belief by S that p is "safe" iff: S would not believe that p without it being the case that p; or, better, iff: as a matter of fact, though perhaps not as a matter of strict necessity, not easily would S believe that p without it being the case that p.)
> Safety   In order to (be said correctly to) constitute knowledge a belief must be safe (rather than sensitive). (Sosa 1999: 142)

Sosa points out that sensitivity and safety are not equivalent, since subjunctive conditionals do not contrapose. Safety is a modal principle. Therefore, one can formulate it by using the notion of possible worlds. Pritchard (2007: 81) formulates the safety principle as following:

> (SP) *S*'s belief is safe *iff* in most near-by possible worlds in which *S* continues to form her belief about the target proposition in the same way as in the actual world the belief continues to be true.[15]

---

[14] Some argue that the fact that sensitivity excludes knowledge-closure shows that sensitivity has to be false. See for example Williamson (2000). For influential criticisms see Vogel (1987) and Sosa (1999). For recent criticisms of sensitivity accounts of knowledge see Melchior (2014a, 2014b and 2015).

[15] For a similar formulation see Pritchard (2005: 156). Pritchard (2007: 283) also considers strengthening the safety principle by demanding that the agent's belief has to be true not just in most of the relevant nearby possible worlds, but in *nearly all (if not all)* of them. Fur a further variant of the safety principle see Pritchard (2007: 292). Sosa (2007) later replaced his initial concept of safety by basis-relative safety which relativizes safety to a belief forming method as Pritchard's principle (SP) does. Pritchard's (2005 and 2007) starting point of his anti-luck epistemology is

What about the safety principle and the case of the scientist? Her beliefs are only based on her intuitions which do not stand in any causal relation to the part of the actual world that makes them true. Hence, it is easily possible that she still has her beliefs on the basis of her intuitions, although her beliefs are false. It is not the case that the scientist would believe that the particles exist, only if they did exist; as a matter of fact, she would easily believe this without it being the case. Therefore, the scientist's belief is not safe according to Sosa's formulation. But since a belief must be safe in order to constitute knowledge, she does not know. The scientist also violates Pritchard's safety principle (SP), since not in most nearby possible worlds in which she continues to form her beliefs based on her intuitions her beliefs continue to be true. Pritchard (2007: 289 and 2009a) explicitly states that safety is necessary for knowledge. Therefore, he is forced to admit, as Sosa is, that the obsessed scientist does not know and that Ci_1 and Ci_2 hold.[16]

What about the contrasting person schoolboy B? Let's take the safety principle (SP) as example. There are possible worlds, where B continues to form his beliefs the same way as in the actual world, but his beliefs are false. These can be possible worlds where O-particles do not exist, but one of the textbook authors mistakenly reports that they do. Whether B knows in the actual world depends on whether such possible worlds are nearby. If they are nearby and if there are sufficiently many of them, then B's belief is not safe according to (SP) and he fails to have knowledge. However, if one regards these worlds as nearby, then knowledge acquisition by reading the textbooks is excluded in general. However, excluding this kind of testimonial knowledge in general seems an unwanted consequence of safety. Hence, those defending safety as the core principle of knowledge face a dilemma: they either must deny testimonial knowledge or accept that schoolboy B knows and that, therefore, the counterintuitive claims Ci3 and Ci4 are true as well.[17]

the common sense claim that knowledge excludes luck. His conception of non-lucky beliefs is closely related to the safety principle, i.e. a belief is non-lucky iff it is safe.

[16] One could argue that any possible world where O-particles do not exist is already far off, since its physical consistency is different from the one of the actual world. In this case, we can take the case of inspector X into account, (or other cases of knowledge by obsession) where this kind of problem does not arise.

[17] Interestingly, Pritchard (2007: 279) also considers the case of a prejudiced detective. In accordance with his anti-luck epistemology he explicitly claims that the detective does not know: "Suppose, for instance, that it was only a matter of luck that the detective stumbled across the crucial piece of evidence which proves the defendant's guilt. So long as her resultant true belief in the defendant's guilt is not lucky, then this poses no problem for the claim that she knows what she believes. In contrast, suppose her belief was only luckily true—suppose, for example, that her belief was based on prejudice rather than evidence, but was true nonetheless—then this would be inconsistent with her possessing knowledge in this regard." Pritchard uses his example for pointing out that his main worry is luck in the truth of the relevant belief and not the luck in coming across evidence.

3.4. *Williamson on reliability and safety*

Williamson regards reliability as a necessary condition for knowledge. "No reason has emerged to doubt the intuitive claim that reliability is necessary for knowledge" (Williamson 2000: 100). Williamson (2000: 123) understands reliability and unreliability as modal states like stability and instability. He thinks that safety is the crucial instance of reliability, which is a necessary condition for knowledge.[18] Williamson (2000: 128) explains the relations between knowledge and safety as following:

> Now assume a connection between knowledge and safety from error […] For all cases α and β, if β is close to α and in α one knows that C obtains, then in β one does not falsely believe that C obtains.

However, Williamson famously reverses the orthodox direction of explanation dominant in epistemology. In his "knowledge first" methodology, Williamson (2000: v) takes the simple distinction between knowledge and ignorance as a starting point from which to explain other things, not as something itself to be explained. According to Williamson, we must use our understanding of knowledge for explaining safety and not the other way round. Consequently, Williamson argues that we have to use our understanding of knowledge to determine whether the similarity to a case of error is great enough in a given case to exclude knowledge. Consequently Williamson (2009: 305) suggests that in "many cases, someone with no idea of what knowledge is would be unable to determine whether safety obtained. … One may have to decide whether safety obtains by first deciding whether knowledge obtains, rather than vice versa."

In order to handle the case of professor O and schoolboy B, Williamson cannot rely on defined notions of safety in determining whether they know, since he inverted the direction of explanation. Therefore, his starting point has to be a judgement about whether O and whether B knows. There are four possible cases:

Case (1): O knows and B knows.
Case (2): O knows and B does not know.
Case (3): O does not know and B knows.
Case (4): O does not know and B does not know.

However, each of these four cases is problematic for Williamson's account.

Case (1): This case seems the desired result given our intuitions. Since safety is, according to Williamson, a necessary condition for knowledge, no possible world where O falsely believes is close. Hence, a world which is exactly the same as the actual world except the fact that O had false intuitions is already far off. But such a restrictive concept of closeness implies an extremely loose concept of safety according to

---

[18] See Williamson (2000, 124). For a discussion on Williamson's account of reliability and safety see also Goldman (2009) and Williamson (2009).

which nearly any belief is safe and safety is not a useful criterion for knowledge anymore. In this case, Williamson's claim that safety is necessary for knowledge becomes inoperable.

Case (2): In this case, the same problematic consequences as in case (1) follow from the claim that O knows. Moreover, the claim that B does not know seems ad hoc, unless we abandon any kind of testimonial knowledge in general.

Case (3): In this case Williamson can make use of an plausible account of safety according to which O does not know for the reason that her belief is not safe but B knows and, therefore, has safe beliefs. However, the problem of case (3) is simply that Williamson is committed to accepting all counterintuitive claims Ci1-Ci4.

Case (4): On the one hand, this case is less problematic than case (3) since it only implies Ci_1 and Ci_2, but not the even more counterintuitive claims Ci_3 and Ci_4. On the other hand, it is–as in case (2)–ad hoc to claim that B does not know, if we do not abandon any kind of testimonial knowledge.

At first sight, Williamson might seem to be in a better position than those who define knowledge in terms of reliability, sensitivity or safety because he need not deny that the detective and the scientist know. But at a closer look, his alternatives are not less problematic.

3.5. *Knowledge without virtue*

Important variants of virtue epistemology are version of epistemic externalism. While process reliabilism focuses on features of the belief forming process, virtue epistemologists mainly focus on features of the believing person. They claim that knowledge has to be the result of a truth-conductive intellectual virtue. Greco argues that simple process reliabilism is too weak. Greco (2000: 177) suggests replacing process reliabilism by the following position, which he calls "agent reliabilism":

> A belief p has positive epistemic status for a person S just in case S's believing p results from stable and reliable dispositions that make up S's cognitive character.[19]

Sosa (2007: 23) regards knowledge as apt performances. Any performance with an aim can have the AAA structure "accuracy: reaching the aim; adroitness: manifesting skills or competences; and aptness: reaching the aim through the adroitness manifest." Sosa (2007: 23) regards beliefs as performances which fall under this AAA structure. "We can distinguish between a belief's accuracy, i.e., its truth; its adroitness, i.e., its manifesting epistemic virtue or competence; and its aptness, i.e., its being true because competent." Sosa distinguishes between animal knowledge and reflective knowledge, a more demanding higher-level knowledge. Leaving the word "knows" undefined, Sosa (2007: 24)

---

[19] The same definition can be found in Greco (1999: 287–88). For Greco's more recent formulation see also Greco (2010).

formulates the core idea of his virtue epistemology as follows.

(a) affirm that knowledge entails belief;

(b) understand "animal" knowledge as requiring apt belief *without* requiring *defensibly* apt belief, i.e., apt belief that the subject aptly believes to be apt, and whose aptness the subject can therefore defend against relevant skeptical doubts; and

(c) understand "reflective" knowledge as requiring not only apt belief *also* defensibly apt belief.

One can easily argue that the obsessed scientist fails to know by adopting one of these virtue epistemologist approaches. It has been assumed that the scientist's intuitions are unreliable belief forming processes and, therefore, not reliable dispositions of her character as Greco's agent reliabilism demands. Therefore, she does not know according to Greco. If one accepts Sosa's virtue epistemology one must admit that the beliefs are accurate because true. It might be subject to debate whether they manifest any epistemic virtue or competence and whether they are, therefore adroit. However, their accurateness does not manifest their adroitness and therefore, they fail to be apt. Hence, Sosa and his followers must confess that O fails to have animal knowledge, and, therefore, also reflective knowledge. Hence, virtue epistemological accounts such as Greco's agent reliabilism or Sosa's virtue epistemology imply that the detective and the scientist do not know and that the counterintuitive claims Ci_1 and Ci_2 are true.[20]

Moreover, one can easily suppose that the beliefs of schoolboy B result from the stable and reliable dispositions to consult serious textbooks and that these dispositions make up his cognitive character and that, therefore, his beliefs have a positive epistemic status as Greco demands. Moreover, his beliefs are according to Sosa accurate because true, they are adroit since they manifest epistemic virtues or competences, and they are apt, since they are true because competent. Hence, B acquires animal knowledge by studying the textbook. Moreover, he can also acquire reflective knowledge if his apt belief is also defensible e.g. against the objection that the textbook is not reliable. Hence, virtue epistemological accounts such as the presented ones not only imply the counterfactual claims Ci_1 and Ci_2, but also the more problematic claims Ci_3 and Ci_4.[21]

To sum up: Inspector X's and professor O's beliefs are neither caused nor causally sustained by a reliable belief forming process. They are insensitive and unsafe and they do not result from an epistemic virtue

[20] Assuming that the obsessed persons form a *second* belief when acquiring evidence fails to be a viable way for the same reasons as for process reliabilists.

[21] Pritchard recently changed his view that safety alone converts true beliefs into knowledge. Pritchard (2009a and 2009b) still defends the necessity of safety to an analysis of knowledge, but he thinks that an ability condition of some sort has to be added. Pritchard (2009a) and (2012) now argues that this *antiluck virtue epistemology* is the right theory of knowledge. However, this account gains the same results with regard to professor O and schoolboy B.

such as stable and reliable dispositions or aptness. Hence, everybody who defends one of these externalist accounts of justification or knowledge is committed to accepting the problematic claims Ci_1 and Ci_2. Moreover, defenders of these externalist accounts are presumably committed to accepting the even more controversial claims Ci_3 and Ci_4 too.

Characteristically, externalist accounts of knowledge and justification do not demand any believer knowledge about the reliability of the belief forming source. Hence, B can acquire knowledge and justified beliefs by simply believing what the textbook tells him and without having any further information about the textbook at all. In all externalist cases, the low standards for knowledge and justification with respect to meta-knowledge conflict with the high standards with respect to the belief forming process that exclude O from knowing.

Notably, problems for externalist knowledge accounts not only arise if Ci_3 and Ci_4 clearly turn out to be false. They already arise if they do not clearly turn out to be *true*. Professor O clearly fails to fulfil any externalist criterion for knowledge and schoolboy B clearly fulfils these criteria. Thus, there should not be any doubts about the truth of Ci_3 and Ci_4 according to externalist knowledge accounts. However, this does not seem to be the case. Rather Ci_3 and Ci_4 are at least controversial borderline cases of knowledge.

Defenders of externalist knowledge accounts could simply argue that their accounts are true and, therefore, Ci_1-Ci_4 are also true and that there is no problem at all. However, this is not a viable strategy, if theories of knowledge shall also *explain* our pre-theoretical understanding of knowledge. Given this is aim, theories of knowledge have to deliver results that resemble our intuitions whether persons know in particular cases and this is not the case, if theories of knowledge clearly imply that Ci_1-Ci_4 are true.

I think one persisting intuition about the cases of X and O is that they do not know until they discovers the evidence, but that they know after having discovered it. However, this intuition cannot be captured by any of these externalist accounts. If X and O clearly know, then these cases of baseless knowledge provide direct counter-examples against externalist knowledge accounts. However, I am not convinced that they clearly know. Rather it seems prima facie disputable whether we should ascribe knowledge to them. However, this *controversy* about the question whether X and O know already confronts externalist knowledge accounts with a problem. X and O clearly do not fulfil any externalist criteria, i.e. their beliefs are to no extend reliably formed or safe and they are clearly not sensitive or apt. Thus, these externalist accounts have it that X or O undoubtedly does not know. They are in no way borderline cases of knowledge according to externalism. However, our intuition seems to be that it is at least subject of discussion whether they know. In this respect, externalist accounts do not adequately capture our intuitions about knowledge.

4. *Lehrer's gypsy lawyer*

The two examples for baseless knowledge presented here are similar to Lehrer's gypsy-lawyer cases. Lehrer (1971 and 1974) presents the case of a gypsy lawyer who proves his client's innocence via a complicated line of reasoning, though his conviction that his client is innocent is completely based on reading the cards. Lehrer concludes that the lawyer knows that his client is innocent, although his belief is in no way caused by his evidence.[22] Lehrer argues against any causal interpretation of the basing relation. Lehrer (1990) claims that the reason a person has for believing something must not be confused with the cause of her believing it. He calls this confusion the *causal fallacy*.

There are similarities and differences between Lehrer's cases of the gypsy lawyer and the racist on the one hand and the case of the obsessed scientist on the other hand. In each of the cases the evidence in no way explains, causes or causally sustains the beliefs. So much for the similarities. However, Lehrer's examples and the cases of obsessed knowledge diverge in important respects. First, we tend to evaluate the causes of holding these beliefs differently. Reading the cards and being a racist are elements of superstition or prejudices which we regard as the opposite of enlightenment and knowledge. At least in our cultural context we tend to have a negative attitude towards superstition and racial prejudices and a positive one towards enlightenment and knowledge. The case of intuitions, in contrast, is less clear. In certain contexts such as scientific discoveries we tend to evaluate the epistemic status of intuitions someway positively, by saying that someone was kissed, touched or inspired by the muse or had a divine inspiration, even if we admit that intuitions are not a reliable guide to truth. From this point of view, one can say that O had a genius moment, when she first came to believe that particles explain the ϕ-phenomenon, although her intuitions did not produce true beliefs in other cases. Being superstitious as the gypsy lawyer or being generally prejudiced as the racist is inconsistent with having knowledge in a way that being inspired by the muse or having a divine inspiration is not. I do not claim that phenomena such as inspiration already constitute instances of knowledge. I only argue that they do not prima facie rule out knowledge.

The second distinction between Lehrer's examples and the case of the obsessed scientist concerns the inter-personal and, hence, social aspects. O not only proves her convictions for her own concerns, she also proves it to *others*. Hence, it is not only her merit that she knows, but also her merit that others know. Moreover, O might have a much better understanding of the evidence and how it is related to the proven theories than those persons to whom she proves it. By only focusing O's beliefs and on Ci_1 and Ci_2, we might just create new versions of

---

[22] Lehrer (1990) presents a similar example of a racist, who has scientific evidence that only members of some race are susceptible to some disease, but who believes this only because of his racial prejudices.

gypsy-lawyer cases, but by taking contrasting persons and Ci_3 and Ci_4 into account, the case against the basing relation receives new support. Lehrer's own cases could have been reconstructed along these lines, but to the best of my knowledge, this hasn't been done yet.[23]

Third, Lehrer wants to argue against externalist knowledge accounts by arguing that baseless knowledge exists. I take a weaker position concerning the existence of baseless knowledge. I only claim that it is controversial whether X and O know. However, this weaker position already suffices for pointing out that pure externalist knowledge accounts do not always capture our intuitions about knowledge.

5. *Overdetermination and pseudo-overdetermination*

There are two possible lines of objection against the claim that O possesses baseless knowledge; first, that she does not possess knowledge at all. All those who defend one of the externalist knowledge accounts sketched above take this line.[24] They are committed to accepting Ci_1 and Ci_2 and typically committed to accepting Ci_3 and Ci_4 as well. Taking the second line of objection means to argue that O possesses knowledge, but that it is not baseless. In this case, one has to argue that even in the cases of O there *is* a kind of causal relation between the beliefs and the evidence that justifies the belief. One way of defending this claim is to argue that O's belief is causally overdetermined because it is caused or causally sustained by her intuitions *and* her evidence. However, the case of O is constructed in a way that her beliefs are only caused and causally sustained by her intuitions. Even after finding evidence, her beliefs are in no respect caused or causally sustained by this evidence. In this respect, her beliefs are not causally overdetermined.[25]

Swain (1981) argues that the belief of the gypsy lawyer and consequently also the belief of O are *pseudo-overdetermined*. He claims in accordance with Lehrer that the gypsy lawyer knows that his client is innocent, but against Lehrer, he argues that the lawyer's belief in the innocence of his client is still based on the complicated line of reasoning. Swain argues that the lawyer's belief is pseudo-overdetermined by the line of reasoning and, hence, based on it because the reasoning would

[23] Interestingly, in his example of the gypsy lawyer Lehrer (1974: 124) denies that the lawyer convinces others by demonstrating his justifying line of evidence, and assumes that the others, "impressed by the similarity of the crimes and eager to believe that the agent of them all has been apprehended, refuse to accept the lawyer's cogent reasoning."

[24] This line is also the most popular reaction to Lehrer's gypsy lawyer case. See for example Harman (1973), Pollock (1986) or Audi (1993).

[25] Goldman (1979) for example notes that beliefs may be over-determined in the sense that they may have a number of distinct ancestral trees. He admits that they need not all be full of reliable or conditionally reliable processes, but at least one ancestral tree must have reliable or conditionally reliable processes throughout. However, the cases of X and O are formulated in a way that there is no ancestral tree that has reliable processes throughout.

have caused the lawyer's belief, if the card reading had not caused it. One can argue analogously that O's belief is pseudo-overdetermined because the evidence resulting from the experiment would have caused O's belief, if her intuition had not caused it.

Swain's account of the basing relation has been criticized by various authors. Kvanvig (1985) argues that we can reformulate the case of the gypsy lawyer in a way that his belief does not fulfil Swain's criterion of pseudo-overdetermination. This can be achieved, for example, by supposing that if the card reading did not cause his belief that his client is innocent, then he would consult a fortune teller, and the evidence would still not cause his belief. In this case, the gypsy lawyer's is not pseudo-overdetermined and, therefore, he does not know according to Swain. Similar counterfactual scenarios can be formulated for O. For example, one can suppose that O would believe that O-particles exist because of religious faith, if she did not have her intuition. In this case, in the nearest possible worlds where O's belief is not caused by her intuition, it is still not caused by the experiment, and consequently O's belief is not pseudo-overdetermined.[26]

6. *Conclusion*

Most accounts of knowledge, especially externalist accounts, share the view that the appropriateness of the way that a true belief is caused or causally sustained is a necessary condition for knowledge. These accounts are committed to accepting that the obsessed scientist and the obsessed detective clearly fail to know. This is counter-intuitive if we ascribe to them other epistemic virtues such as full understanding or praiseworthiness for proving what they believe to others. The view that they clearly do not know becomes even more counter-intuitive if we consider contrasting persons who know because of appropriate belief forming processes but who fail to possess further epistemic virtues.

The conclusion is a moderate one: the appropriateness of the belief-forming process can be one aspect of knowledge among others. The obsessed scientist and the obsessed detective are in one respect not ideal epistemic agents, but so are the contrasting persons who possess weaker understanding than the scientist and the detective and who acquire evidence quite accidentally. Appropriate belief forming processes may be necessary, sufficient, or necessary and sufficient for knowledge in many contexts, but it is dubitable whether they are a necessary condition in all contexts. Any externalist knowledge account that regards a correct belief forming method as necessary in all contexts seems too restrictive.

---

[26] Furthermore, Tolliver (1982) argues against Swain that according to his account, a belief that p could cause a belief that q, nevertheless the belief that q could pseudo-overdetermine the belief that p, which is an unacceptable consequence of Swain's account.

References

Armstrong, D. 1973. *Belief, Truth and Knowledge*. Cambridge: Cambridge University Press.

Audi, R. 1993. *The Structure of Justification*. Cambridge: Cambridge University Press.

Bergmann, M. 2006. *Justification without Awareness: A Defense of Epistemic Externalism*. Oxford: Oxford University Press.

Goldman, A. I. 1979. "What Is Justified Belief?" In George Pappas (ed.). *Justification and Knowledge.* Dordrecht: Reidel: 1–23.

Goldman, A. I. 1986. *Epistemology and Cognition*. Cambridge: Harvard University Press.

Goldman, A. I. 2009. "Williamson on Knowledge and Evidence." In Patrick Greenough and Duncan Pritchard (eds.). *Williamson on Knowledge*. Oxford: Oxford University Press: 73–91.

Goldman, A. I. 2011. "Reliabilism", *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/spr2011/entries/reliabilism/>.

Greco, J. 1999. "Agent Reliabilism." *Philosophical Perspectives* 13: 273–296.

Greco, J. 2000. *Putting Skeptics in Their Place*. New York: Cambridge University Press.

Greco, J 2010. *Achieving Knowledge*. Cambridge: Cambridge University Press.

Harman, G. 1973. *Thought*. Princeton: Princeton University Press.

Klein, P. 2012. "What Makes Knowledge the Most Highly Prized Form of True Belief?" In Kelly Becker and Tim Black (eds.). *The Sensitivity Principle in Epistemology*. Cambridge: Cambridge University Press: 152–172.

Kvanvig, J. 1985. "Swain on the Basing Relation." *Analysis* 45: 153–158.

Lackey, J. 2007. "Why we don't Deserve Credit for Everything we Know." *Synthese* 158: 345–361.

Lackey, J. 2008. *Learning from Words: Testimony as a Source of Knowledge.* Oxford: Oxford University Press.

Lackey, J. 2009. "Knowledge and Credit." *Philosophical Studies* 142: 27–42.

Lehrer, K. 1971. "How Reasons Give Us Knowledge, or the Case of the Gypsy Lawyer." *The Journal of Philosophy* 68 (10): 311–313.

Lehrer, K. 1974. *Knowledge*. Oxford: Oxford University Press.

Lehrer, K. 1990. *Theory of Knowledge*. Boulder: Westview.

Melchior, G. 2014a. "A Generality Problem for Bootstrapping and Sensitivity." *Croatian Journal of Philosophy* 14 (40): 31–47.

Melchior, G. 2014b. "Skepticism: The Hard Problem for Indirect Sensitivity Accounts." *Erkenntnis* 79 (1): 45–54.

Melchior, G. 2015. "The Heterogeneity Problem for Sensitivity Accounts." *Episteme* 12 (4): 479–496.

Nozick, R. 1981. *Philosophical Explanations*. Cambridge: Cambridge University Press.

Plantinga, A. 1993. *Warrant and Proper Function*. Oxford: Oxford University Press.

Pollock, J. and Cruz, J. 1986. *Contemporary Theories of Knowledge*. Savage: Rowman and Littlefield.

Pritchard, D. 2005. *Epistemic Luck*. Oxford: Oxford University Press.

Pritchard, D. 2007. "Anti-Luck Epistemology." *Synthese* 158 (3): 277–297.

Pritchard, D. 2009. *Knowledge*. London: Palgrave Macmillan.

Pritchard, D. 2012. "Anti-Luck Virtue Epistemology?" *Journal of Philosophy* 109 (3): 247–279.

Sosa, E. 1999. "How to Defeat Opposition to Moore." *Philosophical Perspectives* 13: 141–154.

Sosa, E. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford: Oxford University Press.

Swain, M. 1981. *Reasons and Knowledge*. Ithaca: Cornell University Press.

Tolliver, J. 1982. "Basing Beliefs on Reasons." *Grazer Philosophische Studien* 15: 149–161.

Turri, J. 2011. "Believing for a Reason." *Erkenntnis* 74: 383–397.

Vogel, J. 1987. "Tracking, Closure and Inductive Knowledge. The Possibility of Knowledge." In S. Luper-Foy (ed.). *Nozick and his Critics*. Totowa: Rowman and Littlefield: 197–215.

Williamson, T. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.

Williamson, T. 2009. "Reply to Alvin Goldman." In Patrick Greenough and Duncan Pritchard (eds.). *Williamson on Knowledge*. Oxford: Oxford University Press: 305–312.

On the Moral Irrelevance of a Global Basic Structure: Prospects for a Satisficing Sufficientarian Theory of Global Justice

ADELIN COSTIN DUMITRU*
National University of Political Studies and Public Administration,
Faculty of Political Science, Bucharest, Romania

Many important criticisms to the possibility of global justice are advanced following one or another operationalization of the Rawlsian concept of a basic structure. The purpose of this paper is twofold: i) to show that the existence of a global basic structure is irrelevant from the standpoint of justice; ii) to set the stage for a cosmopolitan theory of global justice that employs satisficing sufficientarianism as a distributive principle. One of the main contentions is that the institutional-interactional cut in the recent literature should be transcended. That is, the site of justice should be extended to incorporate both the efficiency of discharging one's duties through a just institutional scheme and the moral value of promoting a good state of affairs through one's own efforts. In order to avoid the overdemandingness objection, however, the selected principles of justice ought to belong to the sufficientarian family. Towards the end of the paper I sketch one such theory, satisficing sufficientarianism.

Keywords: Basic structure, cosmopolitanism, overdemandingness objection, site of justice, satisficing sufficientarianism.

Introduction

The concept of basic structure has been at the heart of numerous criticisms of global justice.[1] Theorists reluctant to accept the prospect for global redistribution have often based their arguments on the qualitative difference between relations mediated by national institutions and relations mediated by global institutions. Depending on their operationalization of the basic structure, they claimed that at the global level we do not have high enough levels of coercion (Blake 2001, Nagel 2005, Risse 2005) or cooperation (Sangiovanni 2007,[2] Klosko 2009) perceived as necessary conditions for triggering considerations of justice. Thus, they arrive at the conclusion that there is an important distinction to be made between full-fledged societal justice and the more ambiguous "global justice", undeserving of the name, and which should require either different, less demanding principles of justice (Rawls 1999, Miller 2007: 79), or be understood in terms of humanitarian duties (Nagel 2005). Albeit value skepticism has ceased to be a respectable option, skepticism about global justice continues to pervade modern political philosophy.

The purpose of this paper is twofold: 1) to argue that whether a global basic structure exists is irrelevant from the standpoint of justice and 2) to set the stage for a cosmopolitan theory of global justice that employs satisficing sufficientarianism as its distributive principle (the "satisficing" component refers to what I consider to be two possible justifications of sufficientarianism: that it is rational for individuals to seek outcomes less than optimal and that moral imperatives do not require one to do more than "enough"[3]). Although the main focus of the paper is on the first subject, I regard the second as one of its logical consequences and as an embodiment of what can be achieved in the domain of global justice once the incredulousness-goggles are abandoned.

---

[1] Although the two concepts are analytically distinct, whenever I refer to global justice in this paper I mean *cosmopolitan* global justice. Cosmopolitanism is based on the idea that each human being has equal moral worth, and that we have certain responsibilities towards all human beings *qua* human beings (Brock and Brighouse 2005: 3–4). I will mostly refer to those who oppose the idea that there are global duties of justice as "anticosmopolitans."

[2] Sangiovanni begins from coercion-based statism but proceeds towards a reciprocity-based statism. He maintains from coercion-based statism the idea that to share participation in reproducing the state coercive system puts us in special relations, which is unlike any other relation we have with individuals beyond our borders and the idea that coercion, private law and taxations are important in limiting egalitarian justice demands to the state. However, he states explicitly that coercion plays only an instrumental role in his reciprocity-based internationalism (Sangiovanni 2007: 18).

[3] Although the problem of how to understand this "enough" cannot be tackled in this paper, I believe that it should be an objective assessment at the very least. As it will be seen in Section IV, I actually propose 2 thresholds, which on the one hand can increase the degree of indeterminacy but on the other can better respond to the diverse circumstances encountered in real life.

The sketch of a sufficientarian theory of global justice that I present in the 4th section is also a way of responding to an important criticism that can be raised against my proposal of holding both individuals and institutions responsible for realizing principles of justice.

I start from the premise that there are three plausible scenarios regarding the concept of a global basic structure, which will be dully called A, B and C: that it already exists (A), that it is absent but it is achievable (B), and that a global basic structure could never be enforced (C).[4] In the first section I present five competing operationalizations of the basic structure (the framing, coercion, pervasive impact, cooperation and controlling influence views). I show how accepting any one of these leads to considerations of global justice (under scenarios A and B). In the second section I mainly deal with scenario B, arguing that taking the status quo as normatively demanding would be self-defeating for a theory of justice.[5] This claim can be accommodated within a Rawlsian framework, which specifies a natural duty to establish just institutions where these are absent (Rawls 1971: 334).[6] The third section introduces the problematic scenario C. I argue that the concept of basic structure does not exhaust the realm of justice. Going further than the narrow Rawlsian understanding on institutions that belong to the basic structure, I hold that for a theory of justice individual conduct outside those institutions should matter too. The moral principles that we choose should not be dependent on the existence of a basic structure. The institutional-interactional cut[7] in the recent literature should thus be transcended. I advance a hybrid approach between interactional and institutional conceptions of justice. That is, the site of distributive justice[8] should be extended to incorporate both the efficiency of discharging one's duties through a just institutional scheme and the moral value of promoting a good state of affairs through one's own efforts. Institutional crafting should be done following two distinct *desiderata*: 1) fulfilling the ends of our preferred theory of justice; 2) allowing individuals to pursue their own reasonable conception of the good within that institutional framework. To achieve such a synthesis, I must show that the latter *desideratum* can be endorsed by reasonable

---

[4] The third scenario also caters to the arguments of those rejecting the epistemological value of this concept altogether.

[5] However, see James (2005).

[6] Chor-Tan (2004) and Gilabert (2007) have pursued this strategy. James (2005: 293), rejects this interpretation of the natural duty. According to James, this duty "is clearly meant to guide conduct with respect to existing practices."

[7] For the distinction see Pogge (1992: 50–1).

[8] The site of justice denotes here where the principles of justice ought to apply. Rawls holds that they apply to the basic structure of a society, interactional/moral cosmopolitans that these apply to individuals, Cohen that they apply to "the patterns of benefits and burdens in a society" (Cohen 1997: 12). The scope of justice, on the other hand, refers to the "constituency" of those principles—do they apply to a single society, or globally?

individuals[9] and also that it does not embody unrealistic expectations from real people. One of the implicit objectives in this section is to show that the necessity of global justice is not dependent on a particular view on the site of justice, and that even if we reject the validity of the idea of a basic structure we hold duties of justice that extend to all other individuals.

This aspect will be argued for in the fourth section, where I am concerned with showing that, in order to avoid the trap of imposing supererogatoriness, the selected principles of justice ought to be sufficientarian. This is also where the *differentia specifica* from similar accounts such as Murphy's (1998) lies: the overdemandingness of pursuing egalitarianism as a "supergoal" (Pogge 2000: 161) precludes the pursuit of personal goals and is thus probably unstable on the long term.[10] Global satisficing sufficientarianism avoids this problem,whilst accommodating the necessity of extending the site alongside the scope of justice.

1. *On the ambiguity of the basic structure argument*

What kind of duties do we bear towards other individuals? What grounds these duties and how weighty are them? Are our duties negative, or positive, and how far do they extend? Placing on the table some of the offers on the menu suffices to show that a minimal consensus on the matter is yet to be found. Pogge (2005: 42) argues that, by "upholding a shared institutional order that is unjust" we "foreseeably and avoidably reproduce radical inequality" and thus we are violating the negative duty not to impose harm on others. Stemplowska (2009) argues that our positive duties extend towards everyone, whereas Val-

[9] I borrow the concept of reasonableness from Rawls. He argues that "persons are reasonable when, among equals, they are ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so" (Rawls 2005 [1993]: 49). Reasonable persons also accept the "burdens of judgment", *i.e.* the fact that other people can develop distinct conceptions of the good and can endorse different comprehensive doctrines (Rawls 2005: 61). The burdens of judgment act as a fundamental source for what Rawls calls reasonable pluralism. The two are perceived by Rawls as inextricably linked—"since we cannot eliminate these burdens, pluralism is a permanent feature of a free democratic culture" (Rawls 2001: 36). Of course, pluralism refers here to the diversity of comprehensive doctrines prevailing in a modern society, a feature which is, according to Rawls, to be cherished (Rawls 2001: 37).

[10] It is perhaps important to note that I partially side here with Rawls' concern with the stability of a well-ordered society. One of the arguments from his Theory of justice for the two principles of justice as fairness is their capacity to stimulate the development of the sense of justice of citizens: "When the basic structure of society is publicly known to satisfy its principles for an extended period of time, those subject to these arrangements tend to develop a desire to act in accordance with these principles and to do their part in institutions which exemplify them" (Rawls 1971: 177). G.A. Cohen (1991), who rejects altogether the compromise of ideal theory by looking at incentives, holds the opposite position.

entini (2015) claims that we have to be Samaritans only towards the members of our state.[11] For Valentini, we are bound by duties of justice towards the needy in virtue of their claim to a sphere of sovereignty that would be imperiled but for our intervention. Hers is an argument that sees the emergence of the duties of samaritanism in our positioning both as "a member of society with entitlements against the state, and as a member of the state with obligations to act on its behalf" (2015: 741). Goodin (1988) argues that, at the level of ideal theory, we would be entitled to treat our "fellow countrymen" with partiality. He rejects the existence of so-called special duties, holding that there are only duties that ought to be discharged following a model of "assigned responsibility." At the level of non-ideal theory, nevertheless, states cannot claim that they are fulfilling their general duty when they give priority to their citizens.

It would be difficult to track down all arguments for or against the idea of positive duties towards compatriots or towards everyone. In order to narrow the discussion, I will focus in this article on criticizing those arguments that relate to the idea of a basic structure. As such, the theories presented below belong to the family of *relational* conceptions of justice, according to which the "practice-mediated relations in which individuals stand condition the content, scope and justification of those principles and that social and political institutions fundamentally alter the relations in which individual stand, and hence the principles of distributive justice that are appropriate to them" (Sangiovanni 2009: 5). We should not consider, however, that all relational conceptions employ the basic structure argument. George Klosko, for instance, presents a public goods arguments for what he calls "compatriot preferences." He argues that, in exchange for our obedience to the state's laws, the state provides benefits in the form of public goods. Nothing comparable exists in the interstate realm. This is why there is moral precedence in the relations between citizens of the same country (Klosko 2009: 244–5). For methodological clarity, I stick in the present paper to those relational views that also employ the basic structure argument.

All the operationalizations of the idea of basic structure that are mentioned here have textual justification in Rawls' own writings. Nevertheless, the reader should keep in mind that these are all particular applications, which at times depart drastically from the Rawlsian scaffolding and which sometimes make abstraction of the larger framework within which the concept of basic structure is embedded.[12] As Abiza-

---

[11] Valentini (2015) argues that we are bound by justice-based duties towards our compatriots and by beneficence-based duties towards everyone, with the qualification that her view of social samaritanism holds the "social" to be "a contingent matter […] in an increasingly globalized world, justice-based help may have to extend beyond national borders."

[12] See Rawls (2001: 25): "We start with the organizing idea of society as a fair system of cooperation and then make it more determinate by spelling out what

deh argued, Rawls himself mainly adopted an anticosmopolitan stance for *methodological* reasons in *Theory of Justice* and *Justice as fairness* (Abizadeh 2007: 359).[13] On the other hand, Rawls emphasizes that one of the reasons for which we need not have a global difference principle[14] is that we have to generalize the idea of public reason and tolerate decent hierarchical societies as long as their basic structures respect a (minimal) core of requirements (Rawls 1999: 45, 58–9). By focusing on the basic structure of each society, critics argue, "Rawls has overlooked a fundamental dimension of the equality of peoples by ignoring the fact that the global basic structure can undermine the equality of peoples unless it is regulated by principles of distributive justice" (Buchanan 2000: 709). Thus, we have mixed evidence regarding the weight of the basic structure argument in John Rawls' own anticosmopolitanism. This is one of the reasons why I will present Rawls' case for seeing the basic structure as the primary site of justice and afterwards I will discuss how the basic structure argument has been employed by anticosmopolitans in its various operationalizations. This section responds to the first strategy of argumentation for global justice—by showing that, in the scenario where a global basic structure exists, we ought to apply the same principles we would apply at the level of a single state. That is, the positive duties we hold towards other individuals are duties of justice proper, not of beneficence.

For Rawls, the basic structure is formed of those institutions which distribute primary goods, determining the division of advantages stemming from social cooperation (Rawls 1971: 7). In the early conceptualization of the basic structure, an institution was considered part of it if it secured "just background conditions against which the actions of individuals and associations take place" (Rawls 1977: 160), ensuring what Rawls later calls the background justice (Rawls 2001: 10). Although the principles of justice should not apply directly to small-scale situations, Rawls holds that a just basic structure "constrains, but does not uniquely determine, the suitable principles of local justice." Rawls avoids offering a clear-cut definition of the basic structure,

results when this idea is fully realized (the well-ordered society) and what this idea applies to (the basic structure). We then say how the fair terms of cooperation are specified and explain how the persons engaged in cooperation are to be regarded (as free and equal)", and the argumentation that follows.

[13] See Rawls (2001: 40): "a political relationship is one of persons within the basic structure of society as a structure we enter only by birth and exit only by death (or so we may assume for the moment). Political society is closed, as it were. We do not, and indeed cannot, enter or leave it voluntarily" (my emphasis). Nevertheless, in a brief discussion on the purposes of political philosophy, Rawls seems to go further than in this pragmatic defense of anticosmopolitanism. According to him, one of the roles of justice as fairness is to contribute to how peoples think of their political and social institutions as a whole, and their basic aims and purposes as a society with a history" (2001: 2).

[14] A position which has been previously endorsed by Rawlsians such as Pogge (1989) and Beitz (1979).

arguing that "were we to lay a definition of the basic structure that draws sharp boundaries, not only would we go beyond what that rough idea could reasonably contain but we would also risk wrongly prejudging what more specific or future conditions may call for, thus making justice as fairness unable to adjust to different social circumstances" (Rawls 2001: 11).[15]

According to Rawls, there are in fact two roles played by the basic structure, corresponding to the two principles of justice as fairness: "in one role the basic structure specifies and secures citizens' equal basic liberties and establishes a just constitutional regime. In the other it provides the background institutions of social and economic justice in the form most appropriate to citizens seen as free and equal" (Rawls 2001: 49). Furthermore, Rawls notes that there are two reasons why the basic structure should be considered the subject of justice. The first line of argumentation is related to the necessity of ensuring background justice, which was mentioned above. A "division of labor" occurs, the endpoint being a situation where individuals "are left free to advance their permissible ends within the framework of the basic structure"; had it not been for a just basic structure, contingencies would have affected the distribution of burdens and benefits in society. *With* a just basic structure, on the other hand, whatever state of affairs is reached is considered by all individuals just: "taking the basic structure enables us to regard distributive justice as a case of pure background procedural justice: when everyone follows the publicly recognized rules of cooperation, the particular distribution that results is acceptable as just" (Rawls 2001: 54). The second reason for the basic structure is that it exerts "profound and pervasive influence on the persons who live under its institutions" (Rawls 2001: 55). Furthermore, the basic structure also has as an important purpose the education of citizens "to a conception of themselves as free and equal" (Rawls 2001: 56).

As mentioned, the arguments for the basic structure put forward by Rawls have to be distinguished from *the basic structure argument* against global justice, which holds that in the absence of a global basic structure there can be no obligations of justice towards foreigners. The case for global justice would be strengthened if it could be shown that irrespective of the interpretation of the basic structure, the present interdependence between citizens of distant countries triggers considerations of justice (I am not concerned with the attractiveness of each operationalization of the basic structure, only with whether or not they lead to global duties of justice).

---

[15] There are two possible readings of this claim. One would follow James' interpretation of Rawls as "reasoning from existing practices all along" (2005: 284). The second, which seems to be adopted by Miriam Ronzoni is that the ever-changing social conditions would ineluctably lead to a reevaluation of what institutions belong to the basic structure. See Ronzoni (2009).

a) *The coercion view.*[16] This interpretation of the basic structure holds that the site of justice consists of those institutions that subject persons to autonomy-violating coercion, in Blake's version (Blake 2001: 272), or those institutions that "make demands on the will of their members […] bringing with them exceptional obligations, the positive obligations of justice", in Nagel's version (Nagel 2005: 130). According to Blake, distributive justice is limited to the basic structure of a society because only its constitutive institutions "stand in need of justification through the use of public reason." For him, subjecting co-nationals to more stringent duties does not reflect unequal concern towards one subset of people. The justification for distributive justice at a narrower scope is that the national institutions are qualitatively different, and as such deserve to be justified: "to the insiders, the state says: yes, we coerce you, but we do so in accordance with principles you could not reasonably reject. To outsiders, it says: we do not coerce you, and therefore do not apply our principles of liberal justice to you" (Blake 2001: 287). Nagel takes into account not only one's subjection to coercive institutions, but also the fact that members of society play a Janus-like role, being both the society's subjects and the ones in whose name its authority is exercised. According to him, our participation in a coercive collective enterprise entails a certain "involvement of agency or will that is inseparable from membership in a political society" (Nagel 2005: 128). As Cohen and Sabel put it, Nagel accepts both weak statism and strong statism. Weak statism holds that the existence of a state is both a necessary and a sufficient condition for the emergence of egalitarian considerations. Strong statism entails that the existence of a state is both necessary and sufficient for triggering any norms beyond humanitarianism's moral minimum (J. Cohen and Sabel 2006). Cohen and Sabel compellingly argue that the normative discontinuity thesis endorsed by Nagel does not take into account the current conditions of global politics: economic integration is more intense than it ever used to; supranational institutions begin to have a considerable impact in fields as diverse as labor standards, environment, rights, food safety standards; the rules made in such structures have a great impact on individuals' conduct and welfare; there is an increasing transnational politics of movements and organizations; even when supranational institutions lack coercive powers, they still have the ability to distribute incentives and to impose sanctions. The fact that nowadays even the least integrated country is a member of fourteen organizations is a compulsory proof that the conditions of interdependence and cooperation that have justice-generating implications obtain globally (J. Cohen and Sabel 2006: 166). Abizadeh shows that, on the one hand,

---

[16] Although the case for a coercion-based interpretation of the basic structure is the weakest of all 5, some textual evidence can be found in Rawls. For instance, he specifies that "justice as fairness is a relationship of persons within the basic structure of society…political power being always coercive power, in a constitutional regime it is at the same time the power of free and equal citizens" (Rawls 2001: 40).

Blake's account is morally unappealing: Blake's government would tell to members of other states: "we not only coerce you, but we coerce you without subjecting our ongoing coercion to the constraints of a legal system and the rule of law, and therefore we have no responsibilities of distributive justice to you" (Abizadeh 2007: 355). Furthermore, the empirical premise on which Blake's argument rests, that there is no international coercive legal system, is invalid (Abizadeh 2007: 356). Even if we bite the bullet and accept this flawed operationalization of the basic structure, the coercion view properly interpreted would still not show that a global basic structure does not exist.

b) *The cooperation view* originates in Rawls' interpretation of society as a "cooperative venture for mutual advantage" (Rawls 1971: 84). Sangiovanni acknowledges that, ultimately, human beings are the unit of moral concern, but argues that the same distributive principle cannot apply indiscriminately. The onerous demands imposed by duties of egalitarian justice arise only among citizens of the same state, since the states provide us with most of the goods "necessary for developing and acting on a plan of life" (Sangiovanni 2007: 3–4). The view he upholds, reciprocity-based internationalism, perceives equality as a "relational ideal of reciprocity." At the heart of his argument lies the moral relevance of the aforementioned ability of being able to develop and act on a plan of life. Since this ability is conditioned by the contributions of our fellow citizens and residents in the state, to them we owe obligations of egalitarian reciprocity (Sangiovanni 2007: 19–20). Albeit the place of one's birth is morally arbitrary, what grounds justice is the idea of reciprocity: "others are owed a fair return of what they have given you, just as you are owed a fair return for what you have given others." The fact that citizens of a state are subject to the same laws and social rules that enable them to "sustain their lives as citizens, producers and biological beings are owed a fair return for what those who have benefited from their submission have received" (Sangiovanni 2007: 26).[17] For Sangiovanni, the brute luck-option luck distinction[18] is valid only under certain circumstances: "the special presumption against inequalities [arising from brute luck] only applies among those who share in the maintenance and reproduction of the state" (Sangiovanni 2007: 29). How does Sangiovanni respond to the most significant objection addressed to Rawls' conceptualization of society as a venture for mutual advantage, that it neglects those who are not able-bodied and thus are not contributing parts of society (Young 2006: 95)? He addresses this objection, but gives what would definitely be perceived by feminists as an inadequate response—"they do not have claims deriving from a conception of distributive equality. This does not mean that they have no

---

[17] Interesting enough, the three hypostases of the human being he envisions correspond to the Arendtian distinction between labor, work and action (Arendt 1958).

[18] For the distinction, see Dworkin 2002: 73.

claims in justice. They have claims which derive from their equal moral worth and dignity as human beings" (2007: 31).

Although the cooperation view seems to lead to some counterintuitive implications, such as disregarding the claims of the disabled, I will leave aside this issue and proceed at showing how a case for global justice can be built on the basis of this operationalization of the basic structure. Take the case of Beitz, who is usually considered to hold a cooperation view of the basic structure.[19] He argues that levels so high of interaction characterize today's world order that global redistributive claims are required. On the other hand, in a scenario where two previously self-sufficient societies would begin exchanging apples and pears, this commercial act would not trigger considerations of justice (Beitz 1979: 65–6). Where to situate the threshold over which justice applies? How to determine whether the levels of interaction are sufficiently high?[20] Is cooperation really a necessary and sufficient condition for duties of justice to arise? Or is it just an instrumental condition? One way of settling the matter would be by employing the conceptual instruments provided by relational equality. According to Anderson (1999: 312), the ideal of equality should be embodied in relational egalitarianism, which considers that equality should characterize a type of social relations between people, instead of being a distribution of non-relational goods. Equality entails not a distributive pattern, but reflects the idea that all people are equally moral agents, everyone having the power to develop and exercise moral responsibility, to cooperate with others according to some principles of justice, to shape and fulfill a conception of their good. If one adopts such a stance regarding equality, one can see why a global basic structure exists. How we perceive ourselves depends not only on how our relations with conationals look like, but how we are perceived by others and how we fare in interactions with foreigners. A permanent position as an outsider cannot but have pernicious effects on one's well-being. Proponents of the capability approach have long argued that the social norms can influence how one

---

[19] For instance, this claim is held by Abizadeh (2007). I am not sure if Beitz would agree with this characterization, since he explicitly mentions that cooperation in a social and economic scheme does not suffice to trigger distributive principles of justice and that pervasive impact and coercion are better harbingers that there are distributive requirements (Beitz 1979: 166).

[20] Arash Abizadeh considers that the cooperation theory shows that a global basic structure does not exist, but that its rationale for holding the basic structure as the site of justice represents a plea for ensuring that the "existence condition of justice, social interaction, obtains at the global level as well" (2007: 327–340). He claims that under the cooperation view, one ought to recognize that "a basic structure is not an existence condition of justice, but an instrumental condition of justice" and that cosmopolitans have the more limited task of showing that a global basic structure would be feasible (Abizadeh 2007: 339). Thus, the problem is relegated to one at the second stage of non-ideal theory, where such agency shortcomings and feasibility issues are dealt with (Ypi 2010).

converts *distribuenda* into capabilities (Robeyns 2000).[21] Thus, adopting a more nuanced cooperation view of the basic structure entails that its scope is global.[22]

c) *The pervasive impact view.* Roughly, under the pervasive impact view the institutions that belong to the basic structure are those with a pervasive impact on persons' life chances. This approach also finds textual support in Rawls' *Theory of Justice,* where he notes that the basic structure is taken as the site of justice because "its effects are so profound and present from the start" (Rawls 1971: 7). According to Abizadeh, the standard anticosmopolitan basic structure argument takes the following form under this interpretation:

P1: The scope of justice consists of those persons whose life chances are pervasively impacted by a society's basic structure.

P2: The range of persons whose lives are pervasively impacted by any given existing basic structure is not global in scope.

C: The scope of justice is not global (Abizadeh 2007: 343–4).

According to Abizadeh, and to other writers as well, the second premise is weak.[23] There are numerous international organizations that exert pervasive influence, such as the International Monetary Fund or the World Trade Organization. The purported case against global justice based on interpreting the basic structure in terms of institutions that impact people's lives is the weakest and as such I will not discuss it any further.

d) *The controlling influence view.* To my knowledge, this operationalization of the basic structure has only been endorsed by Hodgson (2012), who noticed that pervasive influence and coercion represent inadequate criteria for specifying what institutions should be considered as part of the basic structure. He identifies being subject to the rules associated with a basic structure with "being born in the middle of a game that one has no choice about playing." As such, the basic structure ought to be just, since it exerts an influence that determines "how a person can exercise her capacity for a conception of the good", specifying "the rules and constraints through which a person has no reasonable choice but to proceed if she is to adopt and pursue a conception of the good" (Hodgson 2012: 314–5). According to Hodgson, the coercion view makes the illegitimate attempt to treat all institutions as if they were similar to the model of criminal law, whereas the controlling influence view "acknowledges coercion as an important concern,

---

[21] For the distinction between distribuenda and the metric/currency of justice, see Gheaus (2016).

[22] There is another reason that I cannot explore here for advocating interpreting the cooperation view in relational egalitarian terms, i.e. its compatibility with sufficientarianism.

[23] Buchanan (2000: 705): "there is a global basic structure […] a set of economic and political institutions that has profound and enduring effects on the distribution of burdens and benefits among peoples and individuals around the world."

yet it also recognizes that controlling influence can be exerted in ways that are not coercive in a narrow sense but that nevertheless raise fundamental concerns from the point of view of a person's ability to set and pursue ends" (Hodgson 2012: 326). Thus, an institution is part of the basic structure if it influences one's ability to set and pursue ends. Hodgson argues that criticisms such as Cohen's, regarding the impact of an egalitarian ethos on the life prospects of individuals, are accommodated within the controlling influence view: "if sufficiently prevalent, such an ethos has the potential to influence the workings of the basic structure at the deepest level, effectively changing the rules and constraints" (Hodgson 2012: 329). However, his view is indeterminate in an important respect, which leaves room for considerations of global justice to arise—what is and what is not part of the basic structure is determined *ex post facto,* by looking at the alternatives that individuals have; nevertheless, what his theory needs is an independent account of what constitute reasonable alternatives (Hodgson 2012: 334). In an increasingly interdependent world, and in the context of a refugee crisis, borders become part of the basic structure, and the border regime, which is a global institution, gains normative relevance. As such, Hodgson's operationalization of the basic structure is easy to reconcile with global justice claims.

e) *Framing.* Julius (2006) accepts the legitimacy of Nagel's idea that there are responsibilities that we incur although they do not originate in our will. What he doesn't agree with, nonetheless, is Nagel's disregard for the actual acceptance of the terms of cooperation. One has to have the real opportunity to exit the coercive system if she disagrees with the terms imposed on her, otherwise there can be no justification for the coercion. Julius' objections have to be placed in his wider account of the basic structure, which he has developed elsewhere (2003). For Julius, coercive institutions become instruments for influencing other people to serve their purposes. His conclusion is straightforward: one should not use other people to her benefit unless the purpose towards which one aims is compatible to a certain degree with those other people's objectives, or they have their own "reasons to want to come about" (2006: 188). This leads to a criterion that has to be satisfied by the institutions comprising the basic structure: they have to be justifiable to every other person which the choosers of the basic structure frame (2003: 334). More specifically, his view of the basic structure holds that "relations of interactive interdependence create the problem of distributive justice because it is only by reason of her entanglement in those relations that a person is required to justify her shaping of others' actions by appeal to a global distribution of goods" (2003: 344). Framing occurs especially at a transnational level, where citizens of rich countries benefit from the plight of those from poor, underdeveloped countries, who provide cheap labor force and raw materials. As such, we are entitled to speak of a global basic structure, which comprises those institutions through which the advantaged frame the disadvantaged.

We thus have 5 different operationalizations of the basic structure. All of them, if properly defined, can show that there is indeed a global basic structure. If that is the case, then even relational theorists should hold that there is a *prima facie* case for global justice. Nonetheless, I argue that we should go even further. In the next section, I intend to show that even if the anticosmopolitan were right and there were no global basic structure, this should not preclude obligations of justice to arise at the global scale.

2. *Is there a duty to create just basic institutions at the global level?*

Under scenario B), there is no global basic structure yet in place. For Rawlsians, this is the end of discussion. Global justice becomes a mere mirage. However, what should be more important is that individuals have the natural duty to establish the institutions that could provide the resources necessary in order to achieve the ends of global justice. Ignoring this natural duty would be to have an incomplete conception of right (Rawls 1971: 333). One important *caveat:* in this section I do not discuss yet the matter of extending the site of justice. The arguments advanced for establishing a just basic structure even if this were not yet existent usually belong to the institutional family of cosmopolitan theories.[24]

Kok Chor-Tan is adamantly against the idea that global justice can be achieved if we ignore the global institutional context within which countries interact. For him, if we stick with the humanitarian assistance view, we would treat only the symptoms of global poverty, leaving unchanged the structural causes. This is why we need better principles and institutions to regulate the growing interdependency, and "to distribute the burdens and benefits of globalization more equitably" (Chor-Tan 2004: 28–32). For him, a propensity to misinterpret Rawls has been translated into using the notion of the basic structure in order to suppress global justice initiatives. But, if considerations of global justice apply, they must apply regardless of the existing global cooperative scheme. Constraining the applicability of justice to whatever social arrangements we currently happen to have would be an arbitrary bias towards the status quo. If others are vulnerable to our actions or our failures to act, then they fall within the scope of our concern, irrespective of whether there is an established institutional scheme through which to exert our duties (Chor-Tan 2004: 59). Tan affirms that there is a Kantian dimension to this argument, since Kant has noted that considerations of justice come into play the moment our actions have influence on the other. Furthermore, "the degree of global

---

[24] Institutional views apply to institutional schemes, while interactional conceptions "postulate certain fundamental principles of ethics, first-order principles in that they apply directly to the conduct of persons and groups" (Pogge 1992: 48–50).

interdependency is such that even if we were to accept that there are no significant institutions that we are helping to impose on each other, domestic decisions regarding tax laws for businesses, consumption and the deployment of technologies that have environmental implications have potentially grave implications for others beyond the borders of the countries in which these decisions are made. The requirement of reciprocity would demand that such decisions be made only under conditions in which it would be reasonable for outsiders to accept these decisions, even if there were no global institutions mediating the interaction of countries" (Chor-Tan 2004: 173).

What anticosmopolitans do not account for is the fact that "the discrepancy between the density of coercion at the domestic and international levels is not a natural fact about the world, but instead the result of distributive political conflict" (Pevnik 2008: 404). Disregarding the natural duty to establish just institutions at the global level leads to the absence of a just background against which interactions among citizens of different states occur. Pevnik is right to emphasize the fact that confining duties of justice to the national state allows "one's moral status to depend on one's preinstitutional power" (Pevnik 2008: 406). Maintaining the status quo as morally relevant is problematic, since it amounts to what Popper (1947: 60) called "ethical positivism", the inappropriate reduction of norms to facts.

Ypi (2010) remarks that the debates on global justice are vitiated by the fundamental flaw of confusing ideal and non-ideal requirements. Although different contexts mean that the way principles are implemented is different, this occurs at the level of non-ideal theory, which should follow, not ground, ideal theory considerations.[25] Ypi acknowledges that aspects such as coercion, reciprocity, cooperation matter, but they should be of interest only after principles of justice are specified (Ypi 2010: 542). In order for anticosmopolitans' arguments to succeed, they should argue that "there are *no* circumstances of justice at the global level such that [distributive] principles could be required, or that no global relations could warrant a claim for global [distributive] justice" (Ypi 2010: 547).[26] Although she shows the necessity of creating institutions that ensure global basic justice, Ronzoni errs in this way by starting the discussion from non-ideal theory. She claims that empirical studies might be needed in order to ascertain whether the

---

[25] A similar contention can be found in Miklos (2011) where he holds that institutions play a constitutive role in determining the content of principles of justice, i.e. better specify what principles require and how they look like when effectively pursued in non-ideal circumstances (although he does not discuss in terms of ideal/non-ideal theory).

[26] Ypi uses "egalitarian" instead of "distributive." Most debates around global justice focus on the existence of egalitarian obligations of justice. Sangiovanni (2012) constitutes an exception in that he highlights that the distinction between "distributive obligations more demanding than humanitarianism" can also extend to sufficientarian or prioritarian principles.

global order raises problems regarding background justice (2009: 232). Though any practice-dependent account can be susceptible to committing this error, her view presents new arguments for establishing just global institutions. She criticizes the practice-dependent conceptions advanced so far, which erroneously "consider the existence of a basic structure as a necessary condition for some relevant obligations of socioeconomic justice to apply." Arguing for a better understanding of practice-dependence, she stresses the importance of background justice: "a practice-dependent account must also be concerned with social scenarios where full-blown socioeconomic practices with clearly identifiable systems of rules are not in place but where their establishment is required in order to preserve thee justice of other existing practices" (Ronzoni 2009: 234). Although her conclusion is that there should not necessarily be a plain extension of social principles at the global level, she makes an important case *for* the establishment of global institutions that ensure background justice. The views analyzed in the previous sections have disregarded this aspect, leading to some peculiar implications. For instance, the coercive view could not account for the fact that there could be non-coercive types of interaction between citizens that lead to similar consequences as coercive interaction and as such require justification (Pevnik 2008: 407) (such aspects can arise, *inter alia,* from externalities of state actions, or from some apparently consensual interactions which in fact are the result of the lack of existence of background justice).[27] As Barry (1982: 234) argues, relying on the status quo for guiding our considerations of justice would lead to freezing even grotesque allocations of rights.

Thus, if we take the importance of background justice seriously, we ought to strive to establish a just basic structure, even where the actual levels of cooperation/framing/coercion/pervasive impact/controlling influence are not as high in order to trigger in the present considerations of justice. This does not mean that our duties can be discharged only at an institutional level or that we have done our fair share if we contributed to achieving a just basic structure. Institutions matter, but, as I will argue in the next section, so does individual conduct outside of the institutional realm. What matters are not institutions *per se,* but the realization of our principles of justice. We should try, as it were, to achieve what Sen calls *comprehensive outcomes,* which account for consequences, as well as for the "significance of social processes, including the exercise of duties and responsibilities" (Sen 2009: 22). The next section is also of interest for those who reject the epistemic relevance of the concept of basic structure. By advocating an institutional-interactional account, I try to show that the existence of a global basic structure is in fact irrelevant from the standpoint of justice: scenario A has shown that all operationalizations of the basic structure lead to considerations

---

[27] The way the International Monetary Fund has treated the Asian financial crisis of 1997–8 is a well-known example in this sense. See Stiglitz (2002).

of justice; scenario B has argued that we have a natural duty to estab-
lish a global basic structure even if one is not in place for the moment;
in the next section I try to show that realizing this natural duty is in
fact a matter of efficiency and that by themselves institution do not
ground justice; they play only an instrumental role in discharging our
duties. What matters is that, under circumstances where a global basic
structure could not be established, or where institutions fail, our duties
towards others remain intact. The next section thus argues for a hybrid
version of cosmopolitan global justice, which transcends the arbitrary
interactional-institutional cut. The fourth section will show that the
duties mentioned throughout this paper belong to the family of suf-
ficientarian duties of justice.

3. *Revisiting the site of distributive justice*

Throughout the previous sections, several functions of institutions
have been mentioned or implied. Whether they exert a pervasive in-
fluence or controlling influence on individuals, whether they act as a
warranty for pure procedural justice, or if they coerce individuals and
as such require justification, institutions play an important role. But
do they ground duties of justice? More importantly, do they exhaust the
realm of justice?

Nussbaum holds that, although people are the ultimate bearers of
moral duties, we have several reasons for which to uphold an institu-
tional fulfillment of those duties. Collective action issues and the pos-
sibility that others might shirk from their duties lead her to argue for
an institutional route to justice. This way, she argues, individuals are
provided with "broad discretion about how to use their lives...Institu-
tions impose on all, in a fair way, the duty to support the capabilities
of all, up to a minimum threshold. Beyond that, people are free to use
their money, time and other resources as their own conception of the
good dictates. Ethical norms internal to each religious or ethical com-
prehensive doctrine will determine how far each person is ethically re-
sponsible for doing more than what is institutionally required. But the
political task of supporting the capabilities threshold itself is delegated
to institutions" (Nussbaum 2005: 213).

I disagree. According to my conception of the nature of justice, in-
dividuals' duties are not exhausted in institutions. Institutions, de-
spite their advantages,[28] are contingent. Drawing from Ypi's two-stage
theorizing about justice, we could identify two prominent reasons why
justice is not only about institutions. At the level of ideal theory, in-
dividuals have claims on one another prior to the existence of insti-

---

[28] The advantages depend of course on how we define institutions. Basically, the
main advantages of institutions (which can be, *caeteris paribus*, be acknowledged
by all major schools of neoinstitutionalism, i.e. sociological, constructivist, rational
choice or historical), are that they reduce the transaction costs and uncertainty. On
this, see North (1991).

tutions, and these are not deemed irrelevant or magically disappear when institutional schemes are established. At the level of non-ideal theory, there will always be deviations of the institutions from what is required in order to fulfill the ends of our preferred principles of distributive justice. There will always be situations which elude the grasp of institutions.

Let me repurpose one well-known example invoked against (some) prioritarians (and (some) utilitarians), the tyranny of nonaggregation objection. This objection holds that a non-aggregative prioritarian would purportedly let Jones, who is trapped under an electrical equipment in the transmitter room of the World Cup, suffer, as long as rescuing him would bring no matter how trifle disadvantages to a very high number of spectators (Tungodden and Fleuerbaey 2007: 2). Individuals are not to wait for some institution to come in and solve the problem. They have to act, and this is a duty of justice, which can be settled differently according to various principles of justice.

Some, especially Rawlsians, would claim that this is not a matter of distributive justice, but of allocative justice (Rawls 1971: 88). Others would hold that such an example falls under duties of rescue or beneficence, but not under duties of justice. I do not deny the existence of these duties—I make the more limited claim that some apparent examples which purportedly trigger duties of rescue in fact should trigger considerations of justice. Suppose through public debate it is decided that sufficiency is an appropriate moral ideal and that capabilities are selected as the appropriate currency of justice. Suppose further that an institutional framework is established which aims at ensuring that each individual reaches a threshold of capabilities. The system is working smoothly, and most of the previously disadvantaged individuals are brought above the sufficiency threshold. Assume that Cassie is a highly successful businesswoman. She passes one day near Morland, who has recently lost his house.[29] Coincidentally, both Cassie and Morland suffer from a rare condition, which makes them forget at times important pieces of information. For instance, there exists an institution that offers temporary shelter to homeless persons, but neither of them recalls this particular information at the moment. If individuals did not have extra-institutional duties of justice, she could throw Morland some money and leave. However, if individuals have duties of justice, she ought to help him find a shelter, or finance him until he gets a job. This is not merely a duty of charity (which would have been fulfilled even if she had given him a small amount of money), but one of justice—ensuring that each individual reaches a relevant capability threshold. I further argue that she should be indifferent towards Morland's nationality. Of course, sufficiency has as an advantage the fact that it is concerned with absolute deprivation, not with relative standings of individuals. Anticosmopolitans usually endorse sufficien-

---

[29] For simplicity, I won't delve into the problem of responsibility here.

tarian distributive justice at the global level, though they ask for more demanding distributive principles at the level of the nation state (Sangiovanni 2007). This might count as an independent reason for endorsing a sufficientarian conception of global justice, though I won't explore this strategy here.

One more plausible interpretation of the role of institutions in a theory of justice is that offered by Andras Miklos. According to him, institutions constitutively determine the principles of justice. Political institutions "fill out" the substance of a theory of justice "by translating abstract principles of justice into specific rights and obligations for individuals by way of law-making and policy-making" (Miklos 2011: 169). Take rules governing property, he says, and the rule of the lack of legitimacy of transfers of property made under duress. What duress means, nonetheless, has to be interpreted by institutions: "in abstraction from existing institutions, the rule cannot have sufficient specificity" and it would not be easily generalizable. This is one of those instances where individual actions cannot, in the context of social interaction, settle the matter. Institutions that determine the content of justice are necessary here (Miklos 2011: 170).[30] Furthermore, since there are numerous ways that the outcomes suggested by a theory of justice could be obtained, it is up to political institutions to "determine a unique set of rules and provide assurance that they will be adhered to" (Miklos 2011: 173). This means that institutions play a *constitutive role* in theories of justice, determining how the principles of justice are translated in non-ideal contexts, "making the otherwise indeterminate requirements of justice sufficiently determinate by subjecting individual judgment to rules or directives" (Miklos 2011: 175).

It is important to note that Miklos' view on institutions does not represent a criticism to my argument that there are extra-institutional duties of justice. He explicitly mentions that he makes the more limited claim that "principles of justice do not yield a sufficiently determinate answer in the absence of working institutions" (Miklos 2011: 177). I agree, since this is a problem often encountered in non-ideal theory. His position is thus one that endorses neither statism nor cosmopolitanism, but upholds the idea that the *existing* nation-states do not necessarily limit the scope of justice. The global institutional scheme could be reformed "so as to become more sensitive to the demands of global justice" (Miklos 2011: 182). Miklos' arguments find strength when one

---

[30] He identifies such an indeterminacy in Rawls' theory when it comes to the rate of just savings, claiming that in the absence of regulations by institutions, there would be no possibility of knowing that rate or "meeting the requirements of justice prior to and independently of these institutions (171). Rawls holds only that "once the difference principle is accepted, it follows that the social minimum is to be set at that point which, taking wages into account, maximizes the expectations of the least advantaged group...each generation having to put aside in each period of time a suitable amount of real capital accumulation" (besides preserving its culture and ensuring the continuity of a just basic structure) (Rawls 1971: 285).

takes a closer look at history. In the period of Reconstruction that followed the end of the American Civil War the conservative Supreme Court interpreted the Fourteenth Amendment in such a way that a century had to pass that the Civil Rights Movement finish the actual Reconstruction. The Slaughterhouse Cases of 1873 distinguished between national and state citizenship, with the Court arguing that the Fourteenth Amendment protected "only those rights that owed their existence to the federal government...The *U.S.* vs. *Cruikshank* decision of 1876 argued that the amendments following the Civil War required that only violations of the blacks by the states were to be condemned at the federal level. Individual violations were a matter of state authority concern. This allowed the Ku Klux Klan to continue its acts of terror: "in the name of federalism, the decision rendered national prosecution of crimes committed against blacks virtually impossible, and gave a green light to acts of terror where local officials either could not or would not enforce the law (Foner 2002). The way legal institutions actually interpret principles is thus important, and Miklos is right that in the non-ideal world this is going to play a significant role in the way justice is really applied.

The shortcomings of institutions also show why we need to postulate individual positive duties. A similar case is made by Stemplowska (2009), who argues that there are *prima facie* reasons to do what our positive duties require, sometimes at the expense of the negative duty not to support harmful institutions. Contrast this with Pogge's narrower insistence that only membership in a common institutional scheme imposes negative duties not to participate in an unjust institutional order. A purist institutional view "pays too much attention to duty-bearers and not enough to entitlement bearers, to the needy, the hungry and the sick" (Caney 2005: 107–114), that is, those below a relevant threshold. This fetishism of confining justice to an institutional framework can have counterintuitive effects (Murphy 1998: 274).

This emphasis on individual positive duties should not diverge attention from the necessity of establishing just institutions, whenever it is possible (and whenever this does not violate other moral values). There is a reason why a great deal of cosmopolitan writings has been institutional. In a recent paper, Scheffler (2014) presents several plausible reasons why global justice ought to pay attention to institutions. One reason is "normative and conceptual", stressing the fact that "justice is concerned with rights, power and the control of resources", not with acts of "individual beneficence." There are instrumental reasons as well, the problems being too complex to be solved individually. The third reason he mentions is "diagnostic", referring to the structural causes of global poverty, which require a "rewriting of fundamental terms of global political and economic interaction, a restructuring of the practices and institutions of the international order."

However, this instrumental defense of institutional cosmopolitanism should not make us ignore what truly matters—improving the

prospects of the worse off.[31] The position that I defend here is that the site of distributive justice ought to be extended so as to reflect both the efficiency of discharging one's duties through a just institutional scheme and the moral value of promoting a good state of affairs through one's own efforts. Institutional crafting should thus be done following two distinct *desiderata*: 1) fulfilling the ends of our preferred theory of justice; 2) allowing individuals to pursue their own reasonable conception of the good within that institutional framework. As Murphy notes, there is not a clear-cut distinction between the domain of morality and normative political theory (1998: 253). The conception that he advocates, monism, holds that the same principles should regulate institutional and individual conduct. Of course, the point of departure, as it will be emphasized in the next section, is that his is an egalitarian approach to justice, whereas I consider that the global distributive principles ought to be sufficientarian. One important distinction that should be introduced here is that between perfect and imperfect duties. I hold that individuals' obligations are imperfect, while institutional obligations can be easily specified as perfect. For instance, if we remain silent when we can do something to help one not fall under a threshold (for instance, take a non-controversial right, like the right to physical integrity), we would transgress our imperfect obligation to provide reasonable help (Sen 2009: 374–5). This allows us to account for the circumstances in which our actions take place. Had the cost to our own safety been too high, for instance, we would not have been morally required to intervene.[32] Does the distinction between perfect/imperfect duties reintroduce the interactional/institutional cut that I have been arguing against?[33] No, because the nature of the duties remains the same, only the intensity with which different agents have to fulfill them differs. The same obligations befall on individuals as on institutions, the only difference being that, when it comes to the former, one should also account for other values, such as the possibility of leading a life of one's choice. Specifying individual duties as imperfect allows individuals to pursue their own conceptions of the good, within limits.

This latter point will prove to be important in the next section, where I focus on the problem of overdemandingness. Murphy writes that the dualist view on the nature of justice (that different principles apply to institutional and individual conduct) is well represented in literature mainly as a consequence of perceiving the monist view as too

[31] With the caveat that we should focus on comprehensive outcomes (as Sen 2009 calls them).

[32] Specifying individual duties as imperfect duties could soften criticisms such as Saladin Meckled-Garcia's (2008: 256–7) who claims that "it is unclear what coherent principle, primary or secondary, describing perfect duties individual agents could follow that continuously adjusts for fairness in distributive consequences." Under my conception, however, individual duties should only supplement, not replace, duties discharged through institutional action.

[33] I thank Kimberly Hutchings for urging me to clarify this aspect.

demanding: "the standard way of thinking about the problem of what are reasonable moral/political demands focuses on the cost or sacrifice imposed on complying agents. It is true that monistic nonideal theory, if it requires people to do as much as they can to promote equality or well-being seems to be extremely demanding—especially in a cosmopolitan version (Murphy 1998: 289). I will argue that Murphy's own view, however, is too demanding—this is why we need to advocate sufficientarian principles. In order to better link the domains of morality and normative political theory, I will also present a particular version of the sufficiency view, *satisficing sufficientarianism*.

4. *Avoiding overdemandingness—towards a satisficing sufficientarian theory of global cosmopolitan justice*

Extending the site alongside the scope of justice exposes my project to the overdemandingness objection—do we ask too much from individuals if we posit both duties to uphold just institutions and to discharge their (imperfect) duties through personal actions? In order to show how this challenge can be met, I will argue in this section that the selected principles of justice ought to belong to the sufficientarian family of distributive principles.

Some methodological clarifications are in order. I will follow Murphy (2000: 12) in holding that a principle is too demanding when it limits the ability of an individual to live a life of her choice. This occurs at the second level of inquiry postulated by Lea Ypi, that of non-ideal theory. At this level, we have to "take seriously into account non-ideal agency, e.g. the coercive power of states or the associative conditions under which ideal principles of justice become feasible and agents are motivated to promote cosmopolitan initiatives" (Ypi 2010: 543). Allen Buchanan puts forward some plausible criteria for how the transition from ideal to non-ideal theory should be done. According to him, an ideal theory should be accessible, feasible and morally accessible. Feasibility is understood here more narrowly, as compatibility with "human capacities generally." Accessibility refers to "the existence of a practical route from where we are now to at least a reasonable approximation of the state of affairs that satisfies the theory's principles", whereas moral accessibility could be linked to the idea of overdemandingness, in that it asks that there are is no unacceptable moral wrongdoing in the transition from our current states of affairs to the postulated ideal (Buchanan 2004: 61–2). I will focus here on moral accessibility, which I loosely interpret as a criterion whose fulfillment would show that a set of principles do not impose unreasonable moral costs on an individual and thus are not overdemanding.[34]

---

[34] The reader might ask why I associate reasonableness (which, as mentioned above, I take to have almost the same understanding as in Rawls' *Political Liberalism*; a slight change is done below) and overdemandingness. Rawls mentions

Let me now track how this objection may arise and whether it can be met. In his criticism of Rawls' lack of concern with the existence of an egalitarian ethos, Cohen mentions that his ideal society could be erroneously regarded as one in which "a person would have to worry about unfortunate people every time he made an economic decision", a feature which would be appalling to liberals (Cohen 1991: 316).[35] Any monist theory encounters this potential counterargument from liberalism. Rawls, for instance, mentions that a plausible political conception of justice has as a feature its applicability to the basic structure. Furthermore, one of the distinctions between a political conception and a comprehensive doctrine is that the latter "belongs to the background culture of civil society…applying to the daily life, and to its many associations" (Rawls 2005: 12–3).[36] To have a set of distributive principles that apply to the individual actions outside the institutional realm is anathema for liberal Rawlsians, which could interpret the requirements as demanding too much from the individuals.

In order to respond to this potential criticism, I have two main strategies. One, admittedly contentious strategy, is to show that Rawls' own account of political liberalism is too burdensome towards residents of well-ordered societies and towards individuals who under a slightly less stringent operationalization of the idea of reasonableness would be included in the legitimation pool.[37] The second is to sketch a theory of global sufficientarianism that is, in my opinion, a better candidate for a monist theory of distributive justice than any form of egalitarianism.

In his *Law of peoples,* Rawls argues that "a main task […] is to specify how far liberal peoples are to tolerate nonliberal peoples. To tolerate means to recognize these societies as equal participating members in good standing of the society of peoples, with certain rights and obligations, including the *duty of civility,* requiring that *they offer other peoples public reasons appropriate to the society of peoples for their actions*" (Rawls 1999: 59). Notice here an important distinction from *Political liberalism,* where the moral duty of civility is considered to be the capacity of reasonable citizens to appeal to the ideal of public reason when engaging in activities which could alter constitutional es-

---

that "reasonable pluralism of [reasonable] comprehensive doctrines is not an unfortunate condition of human life" (2005: 37), whereas Murphy's interpretation of overdemandingness coincides with principles that ask of some individuals to give up on the life of their choice in order to fulfill a (maximizing) theory's ends. I will assume throughout that there is a moral loss only if those life conceptions are themselves reasonable.

[35] The passage that Cohen has in mind belongs to Nagel: "Most people are not generous when asked to give voluntarily, and it is unreasonable to ask that they should be…It is acceptable to compel people to contribute to the support of the indigent by automatic taxation, but unreasonable to insist that in the absence of such a system they ought to contribute voluntarily" (Nagel 1975: 145).

[36] On the possibility of an overlapping consensus on a set of monist principles see Murphy (1998: 255–6).

[37] I've taken the notion of "legitimation pool" from Friedman (2003).

sentials and matters of basic justice (Rawls 2005: 215–7).[38] Simplifying, matters such as who has the rights to vote, who benefits from fair equality of opportunity or what are the basic needs that ought to be met by the government (Rawls 2005: 214),[39] have to be settled by resorting to arguments from a political conception of justice, which is the matter of an overlapping consensus. That political conception constitutes "a common currency of discussion", where citizens tap in for any arguments they put forward (Rawls 2005: 165). Later, Rawls relaxes this latter condition, and allows arguments to come from reasonable comprehensive doctrines, "provided that in due course proper political reasons are presented that are sufficient to support whatever the comprehensive doctrines introduced are said to support" (Rawls 1997: 784). According to Habermas, this Rawlsian proviso imposes heavy cognitive burdens on both reasonable non-religious citizens, who have to take into account the inputs into the public debate of religious citizens, and on reasonable religious citizens, who have to embrace the necessity of translating their arguments from comprehensive into political ones and to do this without "jeopardizing their own doctrine's exclusive claim to truth" (Habermas 2008 [2005]: 137–144). This is one sense in which Rawls' own theory would seem too demanding for the citizens of well-ordered societies, who have to conduct their behavior in certain circumstances according to a duty of civility from which citizens of decent hierarchical societies are exempted.

One could add to this the problem of a certain skewness in the concept of reasonability towards those citizens who already embrace the legitimacy of political organization (Simmons 2001: 151).[40] Indeed, Rawls offers several possibilities to contest the outputs of public debate: civil disobedience, conscientious refusal, witnessing. Rawls mentions that witnessing is a special case of contestation in which, citizens who usually endorse a reasonable political conception of justice, oppose a decision because of their comprehensive doctrines: "while they may think the outcome of a vote on which all reasonable citizens have conscientiously followed public reason to be incorrect, they nevertheless recognize it as legitimate law and accept the obligation not to violate it" (Rawls 1997: 787). By linking reasonableness to the acceptance of fair terms of cooperation among citizens of a closed society, Rawls "builds too much moral content" in the first concept (Simmons 2001: 151). Closer to the arguments put forward in this paper, one could ar-

---

[38] This idea is more clearly stated in Rawls (1997: 768–9).

[39] See Rawls (2001: 44), for the specification of a principle "requiring that basic needs be met, at least insofar as their being met is a necessary condition for citizens to understand and to be able fruitfully to exercise their basic rights and liberties", and which is lexically prior to the first principle.

[40] "This conception of reasonableness should trouble us. It is not obviously unreasonable to prefer solitude and independence to cooperation. It is surely not unreasonable to prefer more limited or less coercive small scales of cooperation to states" (Simmons 2001: 151).

gue that Rawls puts *too little* moral content in reasonableness. Let's assume that Gordon is a cosmopolitan, but also considers interesting Rawls' ideas of accepting fair terms of cooperation with others and the burdens of judgment (an understanding of reasonableness that extends the scope to which the terms of cooperation apply). His position, nevertheless, would be excluded from public argumentation, since it would be considered unreasonable for denying that "political society should be a fair system of social cooperation for mutual benefit" (Quong 2004: 315). In order for his cognitive input to matter, Gordon would have to renounce at his cosmopolitan ideals. Had he been a member of a decent hierarchical society, he would not have been subject to such rigid norms, though his rights would have not been that secure.[41] This is a second sense in which Rawls' own conception is overdemanding, in that it restricts the scope of conceptions of the good that would be reasonable under a not so narrow understanding of the concept.

One could wonder what is the connection between Rawls' legitimacy account and my overall concern with distributive justice. What unifies these two seemingly separate discussions is the overarching criterion of the possibility to lead the life of one's choice. According to such an evaluative dimension, we can judge both how Rawls' arguments drawn from his political liberalism fare and how my own distributive justice sketch of a theory fares. Following Sen, we could say that this overarching criterion corresponds to a prior principle strategy of argumentation (Sen 1979). Valentini (2012) who argues that both liberal conceptions of justice and accounts of legitimacy share the same fundamental commitment to equal respect and as such can be analyzed together, has endorsed a similar strategy.[42]

I acknowledge the fact that at least one version of monist distributive principles would take a form that would contradict the criterion of moral accessibility. Murphy's endorsement of egalitarianism as a "supergoal" that is to be achieved through personal and institutional efforts has been criticized by Pogge on the grounds that it would ignore agent-relative goals. Full equality being an insatiable (Raz 1986: 235–244) ideal, "even under the best conceivable circumstances, a citizen is morally free to attend to what is important to herself if and only insofar as doing so happens to coincide with her pursuit for the supergoal" (Pogge 2000: 161). I agree that pursuing egalitarianism would be too demanding for individuals, leaving no room not only for agent-relative

---

[41] Space limitations do not allow me to discuss this aspect here. Some arguments for my contention can be found in Buchanan (2006) and Macleod (2006).

[42] "Either a set of institutions instantiates equal respect because its rule meets independent distributive standard, or it instantiates equal respect because its rule is democratically validated" (Valentini 2012: 597). This is an important theoretical claim, because Rawls has adamantly distinguished his account of legitimacy from his conception of justice. See Rawls (2005: 241) for how following the precepts of public reason might lead to results contrary to those preferred by a conception of justice.

goals, as Pogge says (since these could still be ensured through imposing some deontological constraints), but also for personal projects. Cohen's conception, especially, with his criticism of Rawls' focus on incentives, presents such a risk that would surely diminish the appeal of a conception of justice (Cohen 1991). On the other hand, less demanding principles of justice could be compatible with a monism conception. Though he does not endorse such a view, Pogge himself mentions that "monism could be more plausible if specified through a less ambitious goal, like the goal that all human beings have access to freedom and resources above some minimal threshold. Moreover, such goals, once reasonably well achieved, would not have the totalitarian implications of more ambitious supergoals and would not make crushing demands on culture and lifestyle, while leaving plenty of room for the pursuit of agent-centered goals. Under such a supergoal, it would also keep infrequent the occasions on which persons are required to violate just institutions" (Pogge 2000: 163).

The particular version of sufficientarianism that I have in mind is meant to alleviate the concerns that a interactional-institutional hybrid conception of global cosmopolitan justice would be overdemanding. *Satisficing sufficientarianism* appeals to two distinct kind of justification: 1) that it is rational for individuals to pursue good enough ends, and not optimality;[43] 2) that the moral imperatives do not require one to do more than enough, *i.e.* that it would be admirable, but supererogatory to perform actions above a certain threshold of value (Brink 2006: 384). The first corresponds to non-ideal theoretical considerations, while the latter to ideal theory.

More than half a century ago, Simon noticed that the classical view on the rationality of the "economic man" had little to do with reality. The classical view held that the economic man benefited from extensive information regarding the context of choice, a well-organized and system preference order, and "a skill in computation that enables him to calculate, or the alternative courses of action that are available to him, which of these will permit him to reach the highest attainable point on his preference scale (Simon 1978: 9). Simon argued that this account of substantive rationality was incomplete, and that we should look at procedural rationality, *i.e.* the "effectiveness, in light of human cognitive powers and limitations, of the procedures used to choose actions" (Simon 1978: 9). Real human beings do not search for all alternatives, they make cognitive errors, the informational inputs to the decision processes are incomplete and many times they select an alternative that is satisfactory, not optimal (Simon 1985: 295). As such, they are boundedly rational and pursue strategies of satisficing. Satisficing, as opposed to maximizing behavior, is not concerned with getting the most utility out of a situation, but with deriving a certain amount of

---

[43] See also Volacu (2017) for a discussion on consequentialist satisficing, bounded rationality and sufficientarianism.

utility: "In a satisficing model, search terminates when the best offer exceeds an aspiration level that itself adjusts gradually to the value of the offers received so far" (Simon 1978, 10). However, in order to understand the distinction between a maximizer and a satisficer, one has to understand *why* one stops the search. That is, maximizing and satisficing have different *stopping rules:* a maximizer stops because the costs of continuing the search exceed the expected benefits, whereas a satisficer stops because the option he settles with is good enough for her (Schmidtz 1992: 446–7).

Michael Slote has adopted the notion of satisficing and employed it in the model of consequentialist ethics. According to Slote, an act is morally right if it has good enough consequences (Slote 1984: 140). Slote holds that the idea of satisficing consequentialism is anticipated by Popper's negative utilitarianism, where Popper emphasizes that "we have a moral duty to minimize suffering and evil, but no general duty to maximize human happiness" (Slote 1984: 152).

Sufficientarian theories hold that what distributive justice requires is that individuals have enough of some currency or metric.[44] According to Casal (2007: 297–8) the sufficiency view comprises two theses: a positive one, emphasizing the moral importance of people living above a threshold, and a negative thesis, denying the relevance of additional distributive requirements. The sufficiency view has been proposed initially as a counterpart to economic egalitarianism by Harry Frankfurt (1987). In Frankfurt's view, the point of morality is that "each should have enough... If everyone had enough, it would be of no moral consequence whether some had more than others" (1987: 21). However, Shields (2012) has recently argued that sufficientarianism could also be specified as a distributive pattern that "shifts" the urgency of responding to individuals' claims after a certain threshold has been reached.

In this paper, I make the limited claim that the principles that ought to apply at a global level have to belong to the sufficientarian family. I propose the satisficing sufficientarian version as a plausible example of how we could think about cosmopolitan justice, but there are many other conceptions that embody the moral value of sufficiency (Benbaji 2005; Benbaji 2006; Huseby 2010; Shields 2012).

Two-threshold satisficing sufficientarianism: It is morally required to reduce the number of people who are below a minimal threshold. Thus, absolute priority is to be given to those below the minimal threshold. The following disclaimer applies: absolute priority is to be given through institutional action. When it comes to individual action, absolute priority is to be given only if the agents who make the allocation are above the maximal threshold. This is the input from satisficing consequentialism. Between thresholds, strong prioritarian considerations apply: if there are sufficiently numerous people, benefiting them is more important than raising an inferior number of people over the superior

---

[44] The term "currency" has first been used by G.A. Cohen (1989).

threshold. This holds only if there is reasonable expectation that they can in the future be raised over the threshold, taking into account the scarcity of resources and a just savings principle which stipulates that future generations ought to have the opportunity to be above the minimal threshold. Above the superior threshold, the view remains silent. Throughout, Pareto optimality considerations apply, with two exceptions: the absolute priority given to those below the threshold, which trumps trivial losses of those betwen thresholds and considerable losses of those above the maximal threshold; second, a sufficiently large number of people with the prospect of being raised over the maximal threshold trumps a trivial loss of those above the maximal threshold, and, if the number is sufficiently high, considerable losses. I understand the Pareto principle here in a weak sense: a distribution is weakly Pareto superior to another if there is at least one individual better and no individual worse off in the former than in the latter. Since good and bad off are given by individuals' positioning in respect to the thresholds, this prevents at all times the possibility of one individual falling below a sufficiency threshold. Applied to individuals, the exceptions hold only when it comes to trivial losses—that is, we ought to discharge our imperfect, individual duties of justice only if we do not incur considerable costs—this comes once again from the satisficing consequentialism view.

This version of sufficientarianism will ensure that our positive duties towards other individuals are discharged differently when it comes to individual and institutional action. The conception remains monist, in that it specifies the same principles both for individual interactions and for institutional actions. However, it incorporates the concerns for the overdemandingness objection, and it accounts for the fact that as individual agents we can have only imperfect obligations of justice (reflecting Ypi's and Miklos' emphasis on agency-related problems at the level of non-ideal theory). I believe that such a sufficiency view can allow individuals to pursue their own conceptions of the good without sacrificing for this the necessity of fulfilling the obligations of justice. As mentioned above, the view is justified at a normative level by Slote's satisficing consequentialism and at a positive level by perceiving individuals as rational satisficers. As it is, the conception highlights the fact that whether or not a basic structure exists is irrelevant from the standpoint of justice- obligations to raise worse off individuals over a threshold of a preferred currency remain even in the absence of an institutional framework. Institutions should be seen as playing an instrumental-constitutive part in a conception of justice, and not as necessary conditions for considerations of justice to arise.[45]

[45] There remains to be seen how my proposal fares in comparison with other accounts of sufficientarian global justice. I cannot pursue this comparison here, but I can point the reader to the (limited) number of writings that employ sufficientarian distributive principles at the global level: Satz (2010), Kuo (2014), Miller (2009), Laborde (2010).

Conclusions

I have argued in this paper that the basic structure is morally irrelevant, *i.e.* that it does not ground considerations of justice. In the first section I argued that a case can be made that under each five competing operationalizations (coercion, framing, cooperation, pervasive impact and controlling influence) the basic structure is global. In the second section I endorsed the Rawlsian natural duty to establish just institutions, in order to show that, even if there is no global justice, this does not entail that we do not have duties of justice towards residents of other countries. In the third section I argued for extending the site of justice to individual conduct outside of the institutional realm, while in the fourth section I tried to show that the view I put forward is not overdemanding. The institutional-interactional hybrid of global cosmopolitan justice that I propose ought to have as a distributive principle a member of the sufficientarian family. In this paper I argued for a particular version of sufficientarianism, *satisficing sufficientarianism,* which embodies a monist conception of justice, but which is discharged differently: 1) by specifying imperfect duties at the level of individual conduct and 2) by establishing perfect duties at the level of the institutional framework. I have also argued that institutional crafting should be done following two distinct *desiderata*: 1) fulfilling the ends of our preferred theory of justice; 2) allowing individuals to pursue their own reasonable conception of the good within that institutional framework.

Some additional concerns remain. By asking individuals to respect the principles of sufficientarian justice, am I not imposing a certain conception of the good? That is, my view could favor only those individuals who already employ a particular conception of the good, which would be a form of satisficing consequentialism.[46] My response would be that this risk is diminished by the fact that both the distributive rule (sufficientarian) and the ethical conception justifying it (satisficing consequentialism) impose just minimal standards, and could easily be supplemented. Remember that satisficing consequentialism considers any action which goes beyond its minimal specifications to be supererogatory. As such, individuals who hold more demanding ethical conceptions, such as maximizing consequentialism, or various forms of virtue ethics, could easily satisfy the minimal requirements imposed by satisficing sufficientarianism. By focusing on comprehensive outcomes, the view is also compatible with deontological constraints. Thus, following Rawls, satisficing consequentialism could be perceived as a "module, an essential constituent part that fits into and can be supported by various reasonable comprehensive doctrines" (Rawls 2005: 12). Neither does my view exhaust the realm of morality—under certain circumstances, we must intervene according to (sufficientarian) principles of justice. Nevertheless, duties of assistance or morality continue to exist outside this more limited framework.

---

[46] I thank Emil Archambault for this provocative question.

One last *caveat* is that I have only sketched the satisficing sufficientarian theory of global cosmopolitan justice. Much more needs to be done in order to have a proper theory of justice. One of the most important challenges will be to specify the currency of justice, a problem aggravated by the existence of two thresholds. For the moment, I can only point the reader to the vast literature on the problem of a threshold in the literature on sufficientarianism (Arneson 2000; Arneson 2006; Widerquist 2010; Casal 2007). Once again, my case against the basic structure argument ought to be distinguished from my arguments for an institutional-interactional hybrid theory of global justice and also from my arguments for a satisficing sufficientarian distributive principle. One could accept one, two or all three parts, but rejecting one of them does not automatically lead to the repudiation of the others.

References

Abizadeh, A. 2007. "Cooperation, pervasive impact and coercion: on the scope (not site) of distributive justice." *Philosophy& Public Affairs* 35 (4): 318–358.

Anderson, E. 1999. "What is the point of equality?" *Ethics* 109 (2): 287–337.

Arendt, A. 1958. *The human condition.* Chicago & London: The University of Chicago Press

Arneson, R. 2000. "Perfectionism and politics." *Ethics* 111 (1): 37–63.

Arneson, R. 2006. "Distributive justice and basic capability equality. 'Good enough' is not good enough." In Alexander Kaufman (ed.). *Capabilities equality. Basic issues and problems.* New York: Routledge: 17–43.

Barry, B. 1982. "Humanity and justice in global perspective." *Nomos* 24: 219–252.

Beitz, C. 1979. *Political theory and international relations.* New Jersey: Princeton University Press.

Benbaji, Y. 2005. "The Doctrine of Sufficiency: A Defence." *Utilitas* 17 (3): 310–332.

Benbaji, Y. 2006. "Sufficiency or Priority?" *European Journal of Philosophy* 14 (3): 327–348.

Blake, M. 2001. "Distributive justice, state coercion and autonomy." *Philosophy& Public Affairs* 30 (3): 257–296.

Brink, D. 2006. "Some forms and limits of consequentialism." In David Copp (ed.). *The Oxford Handbook of Ethical Theory.* Oxford: Oxford University Press, 2006:380–424.

Brock, G. 2009. *Global justice. A cosmopolitan account.* New York: Oxford University Press

Brock, G. and Brighouse, H. 2005. "Introduction." *The political philosophy of cosmopolitanism.* New York: Cambridge University Press: 1–10.

Buchanan, A. 2000., "Rawls' Law of Peoples: Rules for a vanished Westphalian world." *Ethics* 110 (4): 697–721.

Buchanan, A. 2004. *Justice, legitimacy and self-determination. Moral foundations for international law.* New York: Oxford University Press.

Buchanan, A. 2006. "Taking the human out of human rights." In Rex Martin and David Reidy (eds.). *Rawls' Law of Peoples. A realistic utopia?* Malden: Blackwell Publishing: 150–168.

Casal, P. 2007. Why sufficiency is not enough." *Ethics* 117 (2): 296–326.

Chor-Tan, K. 2004. *Justice without borders, Cosmopolitanism, nationalism and patriotism.* Cambridge: Cambridge University Press.

Chor-Tan, K. 2004. *Justice without borders. Cosmopolitanism, nationalism and patriotism*. Cambridge: Cambridge University Press.

Cohen, G. A. 1989. "On the currency of distributive justice." *Ethics* 99 (4): 906–944.

Cohen, G.A. 1991. "Incentives, inequality, and community." *The Tanner Lectures on Human Values,* Stanford University.

Cohen, G.A. 1997. "Where the action is: on the site of distributive justice." *Philosophy & Public Affairs* 26 (1): 3–30.

Cohen, J. and Sabel, C. 2006. "Extra republicam nulla justitia?" *Philosophy & Public Affairs* 34 (2): 147–175.

Dworkin, R. 2002. *Sovereign virtue. The theory and practice of equality.* Cambridge: Harvard University Press.

Foner, E. 2002. *Reconstruction: America's Unfinished Revolution, 1863–1877.* New York: Harper Perennial Modern Classics.

Frankfurt, H. 1987. "Equality as a moral ideal." *Ethics* 98 (1): 21–43.

Friedman, M. 2003. *Autonomy, gender and politics.* New York: Oxford University Press.

Gheaus, A. 2016. "Hikers in flip-flops: luck egalitarianism, democratic equality and the *distribuenda* of justice." *Journal of Applied Philosophy* doi: 10.1111/japp.12198.

Gilabert, P. 2007. "Contractualism and world poverty." *Social Theory and Practice* 33 (2): 277–310.

Goodin, R. 1988. "What is so special about our fellow countrymen?" *Ethics* 98 (4): 663–686.

Habermas, J. 2008. *Between naturalism and religion. Philosophical essays.* Cambridge: Polity Press.

Hodgson, L. P. 2012. "Why the basic structure is basic." *Canadian Journal of Philosophy* 42 (3–4): 303–334.

Huseby, R. 2010. "Sufficiency restated and defended." *The Journal of Political Philosophy* 18 (2): 178–197.

James, A. 2005. "Constructing justice for existing practice: Rawls and the status quo", *Philosophy & Public Affairs* 33 (3): 281–316.

Julius, A. 2003. "Basic structure and the value of equality". *Philosophy & Public Affairs* 31 (4): 321–355.

Julius, A. 2006. "Nagel's Atlas". *Philosophy & Public Affairs* 34 (2): 176–192.

Klosko, G. 2009. "Cosmopolitanism, political obligation and the welfare state". *Political Theory* 37 (2): 243–265.

Kuo, Y. C. 2014. "Global sufficientarianism reconsidered." *The Taiwanese Political Science Review* 18 (1): 181–225.

Laborde, C. 2010. "Republicanism and global justice." *European Journal of Political Theory* 9 (1): 48–69.

Macleod, A. 2006. "Rawls' narrow doctrine of human rights." In Rex Martin and David Reidy (eds.). *Rawls' Law of Peoples. A realistic utopia?* Malden: Blackwell Publishing: 134–149.

Meckled-Garcia, S. 2008. "On the very idea of cosmopolitan justice: constructivism and international agency." *The Journal of Political Philosophy* 16 (3): 245–271.

Miklos, A. 2011. "The basic structure and the principles of justice." *Utilitas* 23 (2): 161–182.

Miller, D. 2007. *National responsibility and global justice.* Oxford: Oxford University Press.

Murphy, L. 1998. "Institutions and the demands of justice." *Philosophy & Public Affairs* 27 (4): 251–291.

Murphy, L. 2000. *Moral demands in nonideal theory.* Oxford: Oxford University Press.

Nagel, T. 1975. "Review: libertarianism without foundations." *The Yale Law Journal* 85 (1): 136–149.

Nagel, T. 2005. "The problem of global justice". *Philosophy & Public Affairs* 33 (2): 113–147.

North, D. 1991. "Institutions." *Journal of Economic Perspectives* 5 (1): 97–112.

Nussbaum, M. 2005. "Beyond the social contract: capabilities and global justice." In Gillian Brock and Harry Brighouse (eds.). *The political philosophy of cosmopolitanism.* New York: Cambridge University Press: 196–218.

Pevnik, R. 2008. "Political coercion and the scope of distributive justice." *Political Studies* 56: 399–413.

Pogge, T. 1989. *Realizing Rawls.* Ithaca: Cornell University Press.

Pogge, T. 2000. "On the site of distributive justice: Reflections on Cohen and Murphy." *Philosophy & Public Affairs* 29 (2): 137–169.

Pogge, T. 2005. "Real world justice." *The Journal of Ethics* 9: 29–53.

Popper, K. 1947. *The open society and its enemies: the spell of Plato.* London: George Routledge & Sons.

Quong, J. 2004. "The rights of unreasonable citizens." *The Journal of Political Philosophy* 12 (3): 314–335.

Rawls, J. 1977. "The basic structure as subject." *American Philosophical Quarterly* 14 (2): 159–165.

Rawls, J. 1971. *A theory of justice.* Cambridge: The Belknap Press of Harvard University Press.

Rawls, J. 1997. "The idea of public reason revisited." *The University of Chicago Law Review* 64 (3): 765–807.

Rawls, J. 1999. *The law of peoples.* Cambridge: The Belknap Press of Harvard University Press.

Rawls, J. 2001. *Justice as fairness. A restatement.* Ed. by Erin Kelly. Cambridge: The Belknap Press of Harvard University Press.

Rawls, J. 2005. *Political liberalism.* West Sussex: Columbia University Press.

Raz, J. 1986. *The morality of freedom.* New York: Oxford University Press.

Risse, M,. 2005. "What we owe to the global poor." *The Journal of Ethics* 9: 81–117.

Ronzoni, M. 2009. "The global order: a case of background injustice? A practice-dependent account." *Philosophy & Public Affairs* 37 (3): 229–256.

Sangiovanni, A. 2007. "Global justice, reciprocity and the state". *Philosophy & Public Affairs* 35 (1): 3–39.

Sangiovanni, A. 2012. "The irrelevance of coercion, imposition and framing to distributive justice". *Philosophy & Public Affairs* 40 (2): 79–110.

Satz, D. 2010. "Ideals of egalitarianism and sufficiency in global justice." *Canadian Journal of Philosophy* 36.

Scheffler, S. 2014. "The idea of global justice: a progress report." *Harvard Review of* Philosophy 20: 17–35.

Schmidtz, D. 1992. "Rationality within reason." *The Journal of Philosophy* 89 (9): 445–466.

Sen, A. 1979. "Equality of what?" The Tanner Lectures on Human Values, Stanford University.

Sen, A. 2009. *The idea of justice.* London: Penguin Books.

Simmons, A. J. 2001. *Justification and legitimacy. Essays on rights and obligations.* New York: Cambridge University Press.

Simon, H. 1978. "Rationality as process and as product of thought." *The American Economic Review* 68 (2): 1–16.

Simon, H. 1985. "Human nature in politics: the dialogue of psychology with political science." *The American Political Science Review* 79 (2): 293–304.

Slote, M. 1984. "Satisficing consequentialism", *Proceedings of the Aristotelian Society, Supplementary Volumes* 58: 139–163.

Stemplowska, Z. 2009. "On the real world duties imposed on us by human rights." *Journal of Social Philosophy* 40 (4): 466–487.

Stiglitz, J. 2002. *Globalization and its discontents.* New York: WW Norton & Company.

Tungodden, B. and Fleuerbaey, M. 2007. "The Tyranny of non-aggregation versus the tyranny of aggregation in social choices: A real dilemma." London School of Economics and Political Science: Centre for Philosophy of Natural and Social Sciences.

Valentini, L. 2012. "Assessing the global order: justice, legitimacy or political justice?" *Critical Review of International Social and Political Philosophy* 15 (5): 593–612.

Valentini, L. 2015. "Social samaritan justice: when and why needy fellow citizens have a right to assistance." *American Political Science Review* 109 (4): 735–749.

Volacu, A. 2017. "Maximization, Slotean satisficing and theories of sufficientarian justice." *Croatian Journal of Philosophy* 17 (1): 73–90.

Widerquist, K. 2010. "How the sufficiency minimum becomes a social maximum." *Utilitas* 22 (4).

Young, I. M. 2006. "Taking the Basic Structure seriously." *Perspectives on politics* 4 (1): 91–7.

Ypi, L. 2010. "On the confusion between ideal and non-ideal in recent debates on global justice." *Political Studies* 58: 536–555.

Book Reviews

José Luis Bermúdez, *Understanding "I": Thought and Language*, Oxford: Oxford University Press, 2017, 176 pp.

José Luis Bermúdez has been working on the notion of self and self-consciousness for the last two decades. In this book, *Understanding "I": Thought and Language*, he is tackling the conceptual notion of self and self-consciousness. In order to fully understand things that Bermúdez is saying in this book and why he is saying them we need to look at some of his earlier works.

Bermúdez wrote four major philosophical books and numerous articles and publications. These books are (in chronological order): *The Paradox of Self-Consciousness* (1997), *Thinking without Words* (2003), *Decision Theory and Rationality* (2009) and *Understanding "I": Thought and Language* (2017).[1] The two "middle books", *Thinking without Words* and *Decision Theory and Rationality*, are not directly relevant to the book in question: *Understanding "I": Thought and Language*. In *Thinking without Words* Bermúdez is trying to uncover the ontological status and syntactic structure of thoughts that prelinguistic creatures posses (prelinguistic infants, early hominids and animals). In *Decision Theory and Rationality* Bermúdez is trying to demonstrate that rationality can not be explain by any form of decision theory. The only other book in which Bermúdez is exploring the notion of self and self-consciousness (directly) is his first book: *The Paradox of Self-Consciousness*. As it was said at the beginning, in order to evaluate and understand his latest work, *Understanding "I": Thought and Language*, we need to spend some time looking at his previous work, namely, *The Paradox of Self-Consciousness*.

The Paradox of Self-Consciousness is a book dedicated to resolving what Bermúdez calls the paradox of self-consciousness. Now, the point of the paradox is that an account of self-consciousness cannot avoid circularity, may this circularity be explanatory or constitutive. In order to give an

---

[1] There are four types od publications that are not included in this categorization. Firstly, books that are classified as introductions (*Philosophy of Psychology: A Contemporary Introduction* (2005), *Cognitive Science: An Introduction to the Science of the Mind* (2013; 2014)), secondly, books which Bermúdez is not a sole author of (*The Body and the Self* (1998)), thirdly, books that are focused on a single author (*Thought, Reference, and Experience: Themes from the Philosophy of Gareth Evans* (2005)), and lastly minor publications like articles and essays that are not full-fledged books. This exclusion enables us to track the author's thoughts on the subject more precisely.

adequate account of self-consciousness we need to analyse our capacity to think what he calls 'I'-thoughts. What are 'I'-thoughts? 'I'-thoughts are a special way (an ability or a capacity) we think about ourselves, involving concepts and descriptions, that we cannot put to work in thinking about other people and things—namely, the ability to apply those concepts and descriptions uniquely to ourselves. The capacity to think 'I'-thoughts is also fundamental in *Understanding "I": Thought and Language*. So, where is the paradox? The paradox consists in the following: In order to analyse the capacity to think 'I'-thoughts we need to analyse the ability to use the first-person pronoun which seems to require analyzing the capacity to think 'I'-thoughts. And here we have circularity. In order to resolve the dreadful paradox Bermúdez is, throughout the entire book, constructing the notion of prelinguistic self-consciousness.

Firstly, Bermúdez rejects what he calls *The Conceptual Requirement Principle*. *The Conceptual Requirement Principle* states that the range of contents which is permissible to attribute to a creature is directly determined by the concepts which that creature possesses. By doing so the author is stipulating the existence of a nonconceptual content. A nonconceptual content is a form of mental content which can be ascribed to a bearer of that content, without that bearer having to possess the concepts required to specify that content.

Secondly, Bermúdez accepts J. J. Gibson's notion of visual perception through his ecological approach. Briefly, he extracts from Gibson's work the notion of self that is based on spatial self-awareness, like the one in navigation, awareness of orientation and trajectory.

Thirdly, Bermúdez is using the notion of somatic proprioception. Somatic proprioception is a form of perception that provides to the perceiver detailed information about the perciver's position, movement, limb disposition, and other bodily properties. For example: information about balance and posture, bodily disposition and muscular fatigue.

Lastly, Bermúdez is arguing, based on the relevant research, for the existence of prelinguistic social self-awareness. Prelinguistic social self-awareness is manifested in phenomena like joint selective visual attention and coordinated joint engagement which we can observe in infants.[2]

Thus, at the end of his book *The Paradox of Self-Consciousness*, Bermúdez has constructed a solid ground for the notion of prelinguistic self-consciousness.

So, what's the point? The point is that in *The Paradox of Self-Consciousness* Bermúdez, in order to solve the paradox, uses the notions of nonconceptual content, somatic proprioception, visual kinaesthetics etc. to explore the notion of prelinguistic self-consciousness that lies beneath the surface of the iceberg (to use Bermúdez's metaphor from the preface of his current book). Now, nineteen years later Bermúdez, in his current book: *Understanding "I": Thought and Language*, returns to investigate the conceptual notion of self and self-consciousness. After discussing the notion of prelinguistic self-consciousness, he now investigates the surface of the iceberg – full-fledged linguistic (conceptual) self-consciousness.

[2] Certainly, the last four paragraphs cannot adequately present the argument put forward by Bermúdez in *The Paradox of Self-Consciousness*. For a clear and full view of Bermúdez's argument, please see *The Paradox of Self-Consciousness* (1997).

Understanding "I": Thought and Language is structured in the follow-
ing manner. The book has seven chapters (excluding the preface) and each
chapter has between three to five subchapters. These seven chapters are:

1. "I": An essential indexical
2. Sense and understanding
3. Frege and Evans on the sense of "I"
4. Privacy, objectivity, symmetry
5. Token-sense and type-sense
6. I": Token-sense and type-sense
7. Explaining immunity to error through misidentification

The structure of *Understanding "I"* is silmilar to the structure of *The Para-
dox* in the following way. In both books, Bermúdez uses different tools, like
constraints, conditions and criteria, in order to converge at the end of the
book in a single exposition. In *The Paradox* it is an argument that solves the
paradox of the title and in *Understanding "I"* it is the set of conditions that
bring about the concept of "I". The structure is different in books like *Think-
ing without Words* and *Decision Theory and Rationality*. In *Thinking* the
main point is in explaining the notion of nonlinguistic thought throughout
ontology, epistemology and philosophy of language and in *Decision Theory*
Bermúdez's goal was to prove that no decision theory can satisfy the neces-
sary conditions for rationality.

Understanding "I" has a solid and clear structure overall, with chapters
nicely complementing each other. The only objection could be the status
of the last chapter: *Explaining immunity to error through misidentifica-
tion.* In chapter six: *"I": Token-sense and type-sense* Bermúdez makes the
main and final point of the book so the last chapter does not close the book
properly. It would probably be better if chapter seven preceded chapter six.
Understanding "I" employs the following methodology. Bermúdez does not
engage in the ontology of selfhood in a straightforward manner. This is true
for *The Paradox* as well. Instead, he is based on the presupposition that
the self is embodied. That statement is widely and thoroughly explained
and defended in *The Paradox*. In *Understanding "I"* Bermúdez takes the
presupposition as face value.[3] Bermúdez's primary focus is on the complex
interrelations between the epistemology of self-consciousness and its func-
tional role. And the way to approach the epistemology and functional role
of self-consciousness is through an account of what it is to understand the
first person pronoun.

In his first chapter, *"I": An essential indexical,* Bermúdez makes three
main points. These points are: the Expressibility principle, Essential in-
dexicality and the ineliminability of "I". The Expressibility principle sim-
ply states that any thinkable thought can in principle be linguistically ex-
pressed without residue or remainder. This is mostly uncontroversial. The
principle is restricted in application to conceptual thoughts. It does not claim
that there are no inexpressible truths, just that if there are inexpressible
truths they are also unthinkable. What the Expressibility principle does

[3] For the clear and full view of the proposition in question see *The Paradox of Self-
Consciousness* (1997), especially chapter six: Somatic Proprioception and the Bodily
Self.

philosophically is the following. It sets up an equivalence between enter-taining "I" thoughts and understanding "I" sentences. The second point that he makes is Essential indexicality. In Bermúdez's words:

> Essential indexicality (agency): An agent will not typically act upon beliefs about herself unless she knows, through some thought that can only be expressed using "I", that she herself is the person those beliefs are about. Essential indexicality (explanation): When explaining an action in terms of the agent's beliefs about herself, at least one of those beliefs must have as its content an "I"-thought, viz. a thought that can only be expressed using "I". (Bermúdez 2017: 9)

Bermúdez claims that "I" thoughts are fundamental in two relevant and distinctive ways. Firstly; "I" thoughts are the ones motivating an agent, which is called *Essential indexicality (agency)*, and secondly; "I" thoughts are used in explaining an action, which is called *Essential indexicality (explanation)*. The point Bermúdez is making is that "I" thoughts are funda-mental because they integrate the agent's beliefs about the world with her own first person perspective on the world.

From here he makes his final point that "I" thoughts are ineliminable and that any explanation of action without an essential indexical is neces-sarily incomplete.

In chapter two, *Sense and understanding*, Bermúdez is looking at two positions regarding the meaning of a sentence. Firstly, we have a Fregean position which states that the semantic value of a name is a concept that mediates between the name and its referent. That concept might be ex-pressed by a definite description and Frege called it a sense.[4] Secondly, we have Russell/Mill's position which states that the semantic value of a name is simply its referent. Here Bermúdez is taking the "middle ground" which he calls the hybrid view. In the hybrid view Bermúdez keeps the notion of sense, but he is talking of *sense as understanding*. In Bermúdez's words:

> The sense of an expression, whether that expression is a proper name, a logical constant, a predicate or a complete sentence, is what a competent language-user understands when he understands that expression. (Bermúdez 2017: 26)

The chapter is structurally very similar to the second chapter of Bermúdez's *Thinking without Words* where he discusses the nature of thought. There Bermúdez is also looking at two clashing positions; the Fregean approach: thoughts as the senses of sentences and the Fodor approach: the language of thought hypothesis.[5]

In chapters three and four (*Frege and Evans on the sense of "I"* and *Pri-vacy, objectivity, symmetry*) Bermúdez sets out an exposition of Frege's no-tion of the sense of "I" and Evans's notion of the sense of "I". Frege breaks the notion of the sense of "I" in two. We have a private sense of "I" and a public sense of "I". A private sense of "I" is a special way in which I am presented to myself, and that sense is special, private and unshareable. A public sense of "I" is a linguistic device used for communication and under-standing, and its shareable.

---

[4] Bermúdez is using the term Fregean sense: The standard view which does not necessarily correspond with something we might call the Fregean sense: The classical view which would definitely have to involve some sort of Platonism.

[5] For the clear and full view of the discussion in question see *Thinking without Words* (2003), especially chapter two: Two Approaches to the Nature of Thought.

There are three key components of Evans's notion of the sense of "I" that Bermúdez has highlighted. Firstly, Evans follows Frege regarding the notion of the sense of "I". Secondly Evans claims that the sense of "I" is private, but objective, which means that "I" thoughts should exist independently from anyone thinking them,. Thirdly, Evans uses *immunity to error through misidentification relative to the first person pronoun* (the IEM property) in his explanation of "I" thoughts.[6] The IEM property is a special kind of property that "I" thoughts have (arguably not all of them) which states that one cannot think an 'I'-thought without knowing that it is in fact about oneself. 'I'-thought cannot fail to identify the bearer of that thought. An example is the following:

1. John Smith thinks: *I am sitting in a chair.*
2. John Smith thinks: *John Smith is sitting in a chair.*

In (1) there is an "I" thought with an IEM property and in (2) there is an "I" thought without an IEM property.[7]

Bermúdez ultimately rejects both accounts: Frege's and Evans's, respectively. Frege's account of "I" thoughts is rejected because "I" thoughts that can tell us about self-consciousness (private sense of "I") are private and unshareable. Evans's account of "I" thoughts is rejected because "I" thoughts (according to Evans) can exist independently from anyone thinking them. So, what does Bermúdez take from Frege and Evans? From Frege he takes the concept of sense (in a manner described in chapter two) and from Evans he takes the IEM property.

Based on his interpretation of Frege and Evans, in chapter four: *Privacy, objectivity, symmetry*, Bermúdez makes (arguably) his most controversial claim of this book: The Symmetry Constraint.

> The Symmetry Constraint: *An account of the sense of "I" must allow tokens of "I" to have the same sense as tokens of other personal pronouns such as "you" in appropriate contexts.* (Bermúdez 2017: 53).

Bermúdez offers three arguments in defence of The Symmetry Constraint: the same-saying argument, the logical argument and the epistemological argument. The same-saying argument states the following. In understanding a sentence one acquires knowledge of what that sentence says. That knowledge can be reported by a sentence that says the same thing as the original sentence. And finally, if one sentence accurately reports another, then we can reasonably assume that they express the same thought. The logical argument states that the possibility of equivalence in sense between first and second person pronouns is required for meaningful disagreement. Example that Bermúdez provides is that when I say "What you claim is false" and you say "What I claim is not false" then I seem to be denying what you are asserting. Lastly, the epistemological argument states that we need communication in order to transmit knowledge. So, in the right circumstances, when I hear you say something gives me a reason to believe

---

[6] For a more detailed account of Bermúdez's thoughts on Evans see *Thought, Reference, and Experience: Themes from the Philosophy of Gareth Evans* (2005).

[7] For additional clarification regarding IEM see James Pryor: *Immunity to error through misidentification* (1999), Simon Prosser and François Recanati: *Immunity to Error through Misidentification: New Essays* (2012).

it and, if true, it counts as knowledge. Then it follows that the content in both cases is the same.

There are several reasons for concern regarding The Symmetry Constraint. Firstly, The Symmetry Constraint is highly context sensitive. Bermúdez seems to be aware of this fact. That is why his definition ends with: *in appropriate contexts*. Nevertheless, it takes a certain amount of explanatory power away from the constraint in question. Secondly, it could be argued that Bermúdez takes Frege's notion of sense too far and that this notion of sense would be unrecognizable to Frege. Lastly, The Symmetry Constraint could be seen as very strange and counter-intuitive. It seems to us that *I* and *you* cannot have the same sense.

In chapter five, *Token-sense and type-sense*, after installing The Symmetry Constraint, Bermúdez takes a necessary course of action: he breaks the notion of sense in two. On the one hand there is a token-sense: what a speaker or hearer understands when they understand a particular utterance of "I" in a particular context and on the other hand there is a type-sense: what allows a speaker to be described as understanding the linguistic expression "I". Or in the context when we are talking about grasping the truth condition of a sentence involving indexicals:

Grasping the type-sense of an indexical requires:

(a) being aware of how the reference of the indexical is determined by the context of utterance (b) knowing in general terms what it would be for the sentence featuring the indexical to be true (without necessarily being able to identify the referent of the indexical)

Grasping the token-sense of an indexical requires: (a) being able to exploit features of the context of utterance to determine the reference of the indexical (b) knowing a specific truth condition (where this requires being able to identify the referents of the indexical) (Bermúdez 2017: 64)

There are two important questions to be asked here. Firstly, why is breaking the notion of sense in two relevant or useful? Secondly, why is this distinction (token-sense/ type-sense) different from other distinctions of the same type (Frege, Perry and Kaplan)? Bermúdez is claiming that we need the distincion in order to put forward a meaningful account of conceptual self-consciousness. He also claims that he is solving more problems than Frege, Perry and Kaplan. We will see why when we take a look at the next chapter.

Chapter six, *"I": Token-sense and type-sense*, is the main chapter of the book. It is here that Bermúdez summons all the parts of previous chapters and presents his notion of conceptual self-consciousness through understanding the sense of "I". In order to have a satisfactory account of the sense of "I" (token-sense and type-sense) we need to satisfy five constraints.

Constraint 1 (Essential Indexicality): Explain the distinctive cognitive role of "I"-thoughts, as reflected in the two principles Essential Indexicality (Agency) and Essential Indexicality (Explanation).

Constraint 2 (Shareability): Allow thoughts containing the sense of "I" to be shareable.

Constraint 3 (Symmetry): Allow tokens of "I" to have the same sense as tokens of other personal pronouns such as "you" in appropriate contexts.

Constraint 4 (Frege's Criterion): Individuate senses in accordance with

Frege's criterion, so that no two token-senses can be the same if it is possible for a rational thinker to take incompatible attitudes to them.

Constraint 5 (Truth Conditions): Accommodate the distinction between knowing in general terms what it would be for the sentence featuring the indexical to be true (without necessarily being able to identify the referent of the indexical) and knowing a specific truth condition (where this requires being able to identify the referents of the indexical). (Bermúdez 2017: 74–75)

Bermúdez claims that his account of the sense of "I" satisfies the five constraints. This is the main reason why his believes that his account of the sense of "I" solves more problems and has a greater explanatory power from the other accounts (Frege, Perry and Kaplan). As said before, *Constraint 3 (Symmetry)* is probably the most controversial of them.

In his last chapter, *Explaining immunity to error through misidentification*, Bermúdez presents his account for the IEM property. According to Bermúdez not all "I" thoughts have the IEM property. He claims that jugments with "I" as subject (which are conceptual) have the IEM property because they are based on identification-free sources (which are nonconceptual). Identification-free sources are: introspection, somatic proprioception/kinesthesis, visual proprioception/kinesthesis and autobiographical memory.

This is quite a bold claim. The point that Bermúdez is making is that the nature of the IEM property does not rest upon the indexical but on the predicate. It is a special way of receiving information (from identification-free sources) that gives rise to a special kind of predicates that makes the IEM property. The question still remainds: Why is it the case that only judgments that contain a special kind of predicates that are based on a special way of receiving information (introspection, somatic proprioception/kinesthesis, visual proprioception/kinesthesis and autobiographical memory) have the IEM property?

Understanding "I": Thought and Language is overall a concisely written and well-structured book. Bermúdez builds his case precisely and methodically. In each chapter, Bermúdez argues a specific case that is later used as a component for the construction of his notion of conceptual self-consciousness. The exception is the final chapter that does not conclude the book properly because it opens a potentially new subject. At the end the question arises: Does Bermúdez's account of conceptual self-consciousness work? The answer will be in a form of a question: Does it solve more problems than it creates? And that is up to the reader to decide.

<div align="right">
DAVID GRČKI
University of Rijeka, Rijeka, Croatia
</div>

Timothy Hinton (ed.), *The Original Position*, Cambridge: Cambridge University Press, 2015, 281 pp.

In *The Original Position* Timothy Hinton gathered twelve papers that deal with the argument presented by John Rawls in his work on justice in political philosophy and ethics. The original position is, for the vast majority of political philosophers, a fundamental issue in Rawls's theory of justice and politics. It has been more than four decades since *A Theory of Justice* (1971) was published and philosophers still debate over various concepts that Rawls brought up in his book, especially about the original position and the sub-concepts it carries. In the preface, Hinton himself states three major reasons for the longevity of the original position argument. The reasons are, in order of appearance: 1. it captures our intuition that there are morally relevant and morally irrelevant considerations in the process of deciding what the principles of justice are and gives a fresh way of thinking about objectivity in political philosophy; 2. it raises a series of questions that are very important for political philosophers; 3. it has triggered the forming of many alternative positions in political philosophy. If you think about it, he writes, there probably hasn't been a conference concerning political philosophy in which the participants didn't mention Rawls since he published *A Theory of Justice*. So we can probably all agree that there is still the need for publishing books that analyse the concept of the original position.

Now that we have that covered, we need to say something about the collection of papers that Hinton edited. In *The Original Position*, twelve authors take different approaches to Rawls's concept by closely examining different aspects of the original position, laying out some criticism and defending or further criticising Rawls's views. It is a very complex subject but in this collection of papers authors present Rawls's arguments in a fairly simple manner before continuing with a complex analysis. This enables a somewhat unexperienced reader to follow the line of argumentation.

I shall now briefly introduce every paper from *The Original Position* and give a few comments on each of them, starting with the introduction. The introduction (by Hinton himself) gives us a quality insight into Rawls's conception of the original position. I shall give a wider overview of the introduction because it explains the basic components and principles of the original position and I won't repeat the explanations of those concepts when they come up (again) in the rest of the book.

After discussing the three reasons for the longevity of the original position argument I mentioned above, Hinton gives an overview of the original position by examining its position in *A Theory of Justice* (hereafter: *TJ*), the core of the argument and the changes Rawls made to the original position after TJ. The function of the original position in TJ is to stand "between our initial basic convictions and the more abstract principles to which inference will be made." The reason for standing in the middle is the fact that Rawls's conception of justice, justice as fairness, relies on our initial beliefs to take us to the principles that make up a constitutional democracy and are superior to utilitarianism. Because of the aforementioned reason Rawls puts the original position into a broader conception called "reflective equilibrium". The role of reflective equilibrium is to bring into coherence our basic convic-

tions. The original position does this by excluding biased convictions and reasons that we have. That is possible because we are behind the veil of ignorance, which means that we know all the general data about human life and different conceptions of a good life held by people but we do not know any specific data about ourselves and our own life. If we take all this into account, Rawls famously says we will choose two following principles of justice:

> The Equal Basic Liberty Principle: Each person is to have an equal right to the most extensive system of equal basic liberties compatible with a similar system of liberty for others. (TJ, p. 250).
> The Second Principle (which comes in two distinct parts, and in which the first part has a lexical priority over the second): (a) The Fair Equality of Opportunity principle: Social and economic inequalities are to be arranged so that they are attached to offices and positions open to all citizens under conditions of fair equality of opportunity; (b) The Difference Principle: Social and economic inequalities are to be arranged so that they work to the greatest expected benefit of the least advantaged group in society. (TJ, p. 83) [5]

In the part of the introduction which deals with the core of the original position argument, Hinton assesses the maximin rule as the basis for deciding what kind of society we want, the rule that tells us to chose an alternative in which the worst outcome is better than the worst outcome of the other alternative. He also states that we use maximin rule under conditions of uncertainty, and that the original position gives us those conditions. Furthermore, Hinton analyses the second argument for the original position and that is "the strains of commitment". In that argument Rawls appeals to the sense of justice that the parties behind the veil of ignorance have. He continues by saying that they will not choose conditions that will prevent them from honouring the agreement they reach in living together in a society. That also gives stability to the political system. Hinton also writes about two important things that relate to the strains of commitment and treating people as free and equal. First, he assesses Rawls's appeal to the social bases of self-respect. Second, he appeals to the Kantian principle in Rawls's social contract by which we must not treat people as mere means, and that designates that we must share goods by appealing to factors that are not arbitrary.

In examining how the original position changed after TJ, Hinton takes two focal points. Firstly, he examines how Rawls shifted the focus of his theory from figuring what each of us believes about justice to identifying the principles of justice that give the best interpretation of modern liberal democracy. Rawls also focused on the reasonable pluralism in democratic society, which means that we must examine why people give different answers to important question and still remain reasonable. Secondly, Hinton examines Rawls's shift to describing the original position as a device of representation. He states that these shifts in focus help Rawls in solving some problems in his theory, most notably to clarify the distinction between the reasonable and the rational, and to scale-down the ambitions of his theory.

Hinton concludes the introduction by giving an overview of the articles that compose the volume. I will not give an overview of this part because I will give my own overview and assessment of the articles.

The first paper in the collection is written by David O. Brink and it deals with the sustainability of justice as fairness as a counter position to classical utilitarianism (moderate claim) and to mixed conceptions of utilitarianism (ambitious claim). He argues that Rawls has a better chance defending his view against traditional utilitarian views and that mixed conceptions pose a harder problem to solve, although Rawls can defend the moderate claim without defending the ambitious one. In the beginning he assesses the concept of justice as fairness itself. He does this by focusing on the two basic principles used by Rawls. The first principle is the equal basic liberties principle. The second principle has two parts: the difference principle and the fair equality of opportunity principle. The hierarchy of the principles is as follows: equal basic liberties, fair equality of opportunity and the difference principle. This is the general conception of justice. Brink then establishes Rawls's hierarchy in the special conception of justice which is (for him) settled in this way: 1. equal basic liberties, 2. fair value of basic liberties, 3. fair equality of opportunity and 4. the difference principle [23]. The general conception is applied in societies below a certain material well-being threshold, and after we get past the threshold we apply the special conception because we aim to increase basic liberties in that situation. Brink also presents the *extraspecial conception* which distributes all the components of the special conception according to the difference principle which leads to a special conception that is more similar to the general conception. He thinks that this might diminish Rawls's arguments against the mixed conceptions. After that, Brink examines the contractual argument for the special conception, while focusing on the diminishing of the marginal utility and Rawls's use of the maximin rule. He focuses on the use of maximin to counter the use of principle of maximizing utility. Rawls states that the decision making behind the veil of ignorance is a process under circumstances of uncertainty. Brink states that this leads to two gaps in the argument. First, Rawls does not necessarily show that maximin is uniquely rational under uncertainty. Second, it is not shown why we have to decide under conditions of uncertainty in the original position. It is also stated that Rawls gives a harder condition concerning the diminishing of marginal utility when he introduces the aforementioned threshold because he insists that people don't care about additional goods after they receive the highest minimum. Concerning the unbearable situation that might occur if we don't use the maximin rule, Brink thinks that this prevents any risk taking, even a justified one and that it poses a problem for Rawls. After that part of the discussion, the author examines the strength of the special conception against the mixed conceptions. Mixed conceptions accept all of Rawls's principles, except the difference principle. Brink thinks that both mixed utilitarianism and sufficitarianism have strong arguments that diminish the strength of Rawls's arguments for the special conception. He concludes the article with the claim that even if Rawls isn't able to show that his form of liberalism is superior to other conceptions of liberalism, that should not diminish the significance of justice as fairness and that we could reshape the argument to create a stronger opposition to mixed conceptions. That concludes a very concise and clear overview of Rawls's efforts to show the superiority of his conception. Brink lucidly finds gaps in Rawls's arguments

but still isn't too hasty to just dismiss them, but rather tries to find ways in which the argument might be reshaped and that is probably the best part of the article.

In the second paper Gerald Gaus and John Thrasher analyse the problem of rational choice and the original position by examining models of Rawls and Harsanyi. The authors defend *The Fundamental Derivation Thesis*: the justification of a principle of justice J derives from the conclusion that, under conditions C, J is the rational choice of chooser(s) P [39]. They state that many authors don't take Rawls's claim that theory of justice is a part of the theory of rational choice. That is the reason for defending the Fundamental Derivation Thesis and setting aside question of consent. Gaus and Thasher begin with an appeal (and also Rawls's appeal) to rational choice as a common touchstone in a society that has conflicting ideas and intuitions about justice. The principles we come to must be identified as chosen from moral point of view and from the point of view of actual rational individuals, while being recognized as principles of justice. Both conditions are essential and give strength to the original position. Authors proceed by agreeing with Alexander's claim of primacy of individual rational choice, and the claim that it makes a more important part of Rawls's argumentation. They also state that rational choice enables individual choices to have an Archimedean point of view in the moral real. The central part of the paper is comprised from the analyses of the evolution of Rawls's original position and Harsanyi's model. In short, the early models of Rawls have the rational choice of individuals as their foundation, and do not use the veil of ignorance but use a form of the maximin rule. In the examination of the middle model, that culminate in *A Theory of Justice*, authors analyse the following changes in Rawls's theory: "the construction of the information sets; (2) the description of the choosers; (3) the more explicit role of maximin as a principle of rational choice: and (4) a switch in the role of maximin from primarily an argument for the egalitarian principle, to what seems to be the main argument in favour of "the difference principle," which is itself introduced in the middle models."[46]. The new setting enables the choosers to have an impartial (Archimedean) point of view, and together with the introduction of primary goods enables us to choose while using the maximin rule. The final model moves away from rational choice and gives a lot more weight to reasonableness. Gaus and Thrasher sum it up by agreeing with Gauthier and his claim that Rawls's models through the years satisfy the recognition requirement more than the identification one. In presenting Harsanyi's models, authors show how his models are different from Rawls's models. They present his efforts to show how his setting of the original position leads to the choice of average utilitarianism, through the axiomatic model and the usage of equiprobability (assigning equal probability to every outcome of the rules we accept in the original position) and the use of extended preferences (imagining ourselves in the position of other people and evaluating the situation from their point of view). Authors proceed to explain why it is likely that Harsanyi's view fails the identification test and how it might fail the recognition test. Gaus and Thrasher conclude that the Archimedean point of view is alien to most people and that Rawls's later conception, as well as Harsanyi's conception, use problematic principles of

rational choice. They believe that Rawls's early model has the best chance of solving this problem. The authors put focus on relatively neglected role of rational choice in Rawls's theory and their claim that contract theory is secondary to rational choice is certainly intriguing. Their analysis of identification and recognition conditions is very useful because it gives us another way of examining if the original position theories really generate principles of justice for our societies.

The third paper is called "The strains of commitment" and is written by Jeremy Waldron. In the paper Waldron argues that the strains of commitment argument is stronger than the arguments for using the maximin rule, and that Rawls should put more emphasis on the responsibility of people that need to honour the agreement after the veil of ignorance has been lifted. He tries to defend the strains of commitment argument against two big objections, the malcontents objection and the model-theoretic objection. The malcontents objection has two parts. Firstly, it states that people might reject Rawls's principles if people reflect on how much better off they would be under alternative principles. Secondly, the objection states that some people could reject Rawls's principles by stating that they would be better off if they weren't living under principles of justice at all. The answer to the first objection is that people should be able to live under principles of justice, and that they can't complain if they are not willing to live under Rawls's principles or only have strong preferences against it. The second part of the objection has a relatively short answer which states that people who are tempted to violate the principles of justice and want to benefit from previous injustice don't need to be taken into account in determining what justice is. The model-theoretic objection says that the strains of commitment argument uses features of application of the principles in society, while determining the principles of justice in the hypothetical model. But Waldron thinks that the base-model objection can be met by a deeper analysis of the liberal orientation of Rawls's theory. He does this by appealing to pluralism and publicity that guide our search for justice and give it a basic grounding in the liberal tradition, and by examining the original position as a heuristic device that helps us shape "model conceptions". He also criticizes Ackerman's attempt to construct an ideal theory with the perfect technology thought experiment. Waldron thinks that perfect technology (though useful in putting aside question of implementing the principles of justice) disregards three important features of liberalism: the state is not only our protector but the greatest threat to our rights (if we give it too much power), preservation of justice is mostly based on voluntary acts of individuals, and we build on those voluntarily acts that enable people to have mutual respect and to cooperate. Waldron concludes that we have the best chance of recognizing the limits of our agreement and capture the spirit of freedom and mutual respect if we use the strains of commitment argument. The paper itself is fairly simple, because it takes one big argument and two big objections against it, but that makes it very easy, understandable, and enjoyable to read. It is straightforward in giving us the answers to important questions of commitment to the principles of justice (once we agree upon them). In defending the strains of commitment Waldron also (quite successfully) defends the contractarian argument put forward by Rawls.

The fourth paper is written by John Christman, and is entitled "Our talents, our histories, ourselves: Nozick on the original position argument." Christmas focuses on Nozick's arguments against the original position and tries to show the reason why his critique fails, while he also tries to find some points of convergence between Nozick's and Rawls's theory. Author begins with a brief overview of Rawls's theory and the original position, focusing on the parts Nozick criticizes, especially the difference principle. He also gives a basic overview of Nozick's use of the Lockean proviso of leaving "enough and as good" for others while appropriating and exchanging goods in a society. From there Christman proceeds to lay out Nozick's criticism of the difference principle and the original position by analysing four lines of criticism that overlap. The first line of argumentation is based on sharing the benefits of social cooperation, and has two parts (both concerning the division of social surplus): considering life outside society for the better favoured, and their marginal contribution to the social product. The first part is based on picturing the better off as Robinson Crusoes that produce on their own island and don't need to share the product of their labour. This is quickly rejected by using Pogges critique of that approach and Rawls's claim that we cannot envision life outside society any more. The second part poses a better challenge to Rawls but still (according to Christman) fails because although it appeals to calculation of better marginal contribution by the better off, which is also needed in the usage of the difference principle, it fails to capture the need for deriving principles of justice (it still doesn't render the veil of ignorance as useless). The second line of criticism is based on bias against historical principles. The argument is based on the fact that the original position rules out some theories of justice, including Nozick's entitlement theory. Christman argues that Nozick uses micro examples to show how this in unfair, while he in fact disregards the purpose of the adopting the original position in a process of getting to unbiased institutions. The third argument is centred in slavery of the talented and the priority of liberty, which also relies of micro examples. Nozick states that taxation of the talented (who earn more) is like slavery and that it interferes with their liberty. Christman replies that policies which prevent them from getting too rich, don't interfere with their conception of the good and their pursuit of their life goals. The last objection is based on the arbitrariness argument, which claims that Rawls also takes into account arbitrary factors and that he doesn't justify deviations from equality. The answer to this criticism is in the appeal to the level of institutions, because "Rawls does not claim that arbitrary contingencies cannot play a role in how people end up in their relative position of social advantage." [90] Rawls claims that we should not take arbitrary contingencies into account when determining just institutions. Christman finishes the paper by giving general reflections on the original position and Nozick's entitlement theory, while comparing the two. He does this in two parts, first examining different meanings of personal entitlements of people and their complexity and then proceeds to briefly analyse justice in the non-ideal world. Christman's paper gives us a clear view of differences between Nozick's and Rawls's theories, as well as differences in their aims. He does a great job in using TJ to present Rawls's view, because Nozick criticizes that part of Rawls's work, but still explains

how Rawls changed his theory through the years, and how that also affects Nozick's critique of the original position. He also gives clear and simple answers to Nozick's arguments, not only through his own comments and secondary literature but also by appealing to Rawls himself.

The fifth paper, written by Matthew Clayton, deals with similarities and differences between hypothetical reasoning in Rawls's theory and Dworkin's theory. He does this mostly through the analyses of the role of hypothetical reason in their accounts of justice (Rawls's justice as fairness and Dworkin's equality of resources). Clayton begins with Dworkin's assessment of the original position, and supports the claim that although Dworkin states that a hypothetical contract isn't as binding as a real contract, he does not reject hypothetical reasoning as a whole but rather states that it is grounded in some deeper principle (for him that is right to equal concern and respect). Rawls's reply is that the original position is situated in a larger framework and its successful integration within a conception of justice, and he rejects the grounding Dworkin proposes. Clayton continues by briefly analysing Dworkin's concept of using thought experiments that do not involve hypothetical reasoning, an auction of resources and its envy test (which insures that no one wants the set of resources of anybody else in the end), which he combines with the hypothetical insurance theme. That includes putting individuals behind a "thin" veil of ignorance in which the people have more information than in Rawls's but still don't know what their chances are of being the one that lacks internal resources. Two objections that also arise are concerned with the difference principle, which according to Dworkin fails to respond to morally relevant differences between individuals and the original position that uses a veil of ignorance that is too thick and that it affects the fairness of a society. The question of excluding their conceptions of the good is problematic for Dworkin, but also for other authors (Clayton uses Nagel's objection), but Rawls tells us that under conditions of reasonable plurality and by virtue of "comprehensive doctrines" that exist, advancing our good might not advance the good of everyone (which is exactly what he wants to achieve). The line of argumentation that occupies Clayton in the remaining part of the text is Dworkin's effort of creating a connection between justice and an individual's values. He focuses on Dworkin's use of the envy test to show the value of using a thin veil of ignorance. The envy test enables us to compare ourselves with others, but we can be entitled to compensation only if we can't satisfy out life goals. In short, if we don't feel envy towards others, than we are equal. This is used to critique Rawls's approach that uses primary goods, and does not take into account that some people don't value primary goods as much. Clayton also explains Dworkin's *ex ante* envy test of insurance buying argument, and the hypothetical insurance market that eliminates information that can produce an unfair outcome. All this also gives "an account of equality that is responsive to people's ambitions for their lives". This is used to compare the thin and thick veil of ignorance, and during the comparison Clayton states that Dworkin's conception may be superior for comparing people's lives and that it makes principles of justice more acceptable to people. Clayton actually gives an excellent overview of challenges that Dworkin's view of liberalism poses to Rawls. Maybe the biggest factor is that Dworkin also wants

to achieve a society which is based on equality and liberalism, but (maybe) gives a better account of how people should compare themselves to others while using Rawls's own device, the veil of ignorance (which he modifies). Clayton captures that fact very clearly and understandably.

The sixth paper "Feminist receptions of the original position" is written by Amy R. Baehr, and in it Baehr evaluates the original position through feminist considered convictions while trying to see if we can have a plausible account of feminism that is also contractarian. She starts by giving ten feminist considered convictions about injustice at the level of social institutions and the society itself (or societies), while also introducing the conception of the gender system that we also find in Rawls's work. Baehr than considers two reasons, brought up by Rawls, the convictions could not be met by his theory. First is "ought implies can" related and is based on fixed features of the social world and the second is that the gender system might be amended only through measures that are ruled out by the principles of justice. She proceeds by taking into account coercive remedies in a well-ordered society. She analyses Rawls's account of just and unjust arrangements while examining the possibility of injustice towards women enduring in a well-ordered society. Baehr then, taking Rawls's instructions, tries to adjust feminist considered convictions to the principles of justice. Violence and discrimination are forbidden by Rawls's principles, but female primary parenting and sexist socialization aren't necessarily recognized as unjust. The possibility (and often sad reality) of women constituting the majority of poor people also isn't counted as unjust because the difference principle doesn't necessarily amend that situation. In short, the original position does not recognize some feminist considered convictions as unjust, and also prevents effective remedies towards others. Baehr also states that Rawls might propose reconciliation with most of the gender system. After that she gives two feminist proposals to modify original position. The first one is based on extending the scope of application of Rawls's principles, a proposition Baehr mostly bases on Susan Okin's feminist full basic structure view. The second proposal is based on an attempt to situate considered conviction about dependency into the initial situation by proposing the addition of not fully cooperating individuals to the situation. In the end, Baehr proposes that we take Rawls's instruction and continue the process of reflective equilibrium. The paper gives an interesting perspective of possibilities of forming liberal conception of feminism through Rawls's theory. While examining feminist considered conceptions and their status in a well-order society Baehr shows that Rawls's theory has problems in accommodating the demand that feminism poses.

David Estlund is the author of the seventh paper, which is entitled "G.A. Cohen's critique of the original position". In the paper, Estlund presents Cohen's critique trough three lines of argumentation: "fact dependent foundations", "justice as regulation", and the claim that the original position gives to morally bad fact. He first present Cohen's relative and ultimate claim of fact-independence. The relative claim tells us that principles are grounded in deeper principles that depend on different facts, while the ultimate claim tells us that those principles are grounded in a principle that does not depend on facts. That forms a base for analysing Cohen's objec-

tion which is based on unearthing (Estlund names the concept) the deeper principles. Estlund first examines the formal objection, which states that Rawls's principles of justice in fact lie on deeper principles and considers a possible counter argument that may be found in Rawls's theory. He continues by analysing the substantive objection, which critiques Rawls's original position as a method that is made for choosing rules of social regulation and not for choosing the principles of justice. Estlund says that the two objections differ in the points of their critique: the formal objection is based on the facts and the substantive one is based on the notion of regulation. He states that the strong distinction between rules of regulation and principles of justice gives strength to Cohen's critique, especially if we closely examine it apart from the formal objection. After briefly appealing to Nozick's and Cohen's objections that Rawls is question begging in the process of deriving the principles of justice, Estlund moves to Cohen's critique that states that the original position incorporates values that have nothing to do with justice. Although he thinks that this objection is far more problematic for Rawls's theory, Estlund states that this objection relies on the problem of social policy. That is why he examines Cohen's objection based on tax brackets and exactness and differential care, arguing that Rawls's theory can be successfully defended from both objections. He concludes by presenting the third line of critique, the one about constructivism being sensitive to morally bad facts, while stating that although it is not an elaborated objection it still poses a serious challenge to the original position. The paper gives us an overview of one of the most elaborated critiques of the original position, which has grown during time. Estlund presents Cohen's view in a simple and understandable manner, just as he announces in his introduction. He gives answers to the critiques by showing how Rawls's view can be modified to accommodate the objections, while also showing how some objections fail because they are inconsistent and question begging. Still he manages to show us that we must take those objections seriously by pointing to some weak spots in the original position.

Timothy Hinton himself wrote the eighth paper in the collection he edited, entitled "Liberals, radicals, and the original position". Hinton aims to show how social theory shapes Rawls's liberal conception and his principles of justice, while also showing how radicals should contest his view. He begins with a brief overview of Rawls's concepts that are important for discussion with the radicals (two moral powers, veil of ignorance, self-respect), while showing that Rawls's work is dominated by a form of liberal legalism and primacy of the liberal freedoms. After that Hinton tries to show that Rawls chooses the parties in the original position to have an ideal-historical approach and work within the thin social theory, without using real history and rich normative or evaluative language. That, according to Hinton, opens Rawls to criticism from radicals. Then he presents an opposite view by presenting a simplified version of racial capitalism, and its factual part about the emergence of white domination with the development of capitalism, as well as its explanatory part which focuses on conventions that whites put in place for the purpose of racially dividing the world in order to dominate the blacks. This theory is explained and then put into the original position framework. Hinton does this to create an argument which

contrasts telic autonomy and individual sovereignty, and also contrasts liberal-democratic specification of the equal basic liberties principle with the radical-democratic specification. He argues that if we enable parties in the original position access to real history they would choose the radical specification, because they want to preserve their self-respect (and would do that more effectively) and because racial domination would prevent a large part of a society to develop telic autonomy (they would have a good position to form their own system of ends). Hinton also briefly discusses problems that arise while Rawls uses thin-ideal social theory to get to the second principle of justice. He finishes the paper with three objections to his argument: the appeal to the Ockham's razor in favour of Rawls's theory, the fact that Rawls presents fourstages of implementing the principles of justice to laws and the claim that Hinton mixes them up and the ideal theory objection that states that Rawls focuses on the ideal and Hinton on the non-ideal cases. Hinton thinks that all these objections can be met as he puts forward his replies. The paper as a whole provides strong basis for arguing that Rawls does not produce principles that can effectively preserve our moral powers in a non-ideal circumstances, mostly because it does not appeal to real historical facts and thick social theory. Hinton manages to capture this problem while also giving a viable alternative using the radical view in the original position framework and shows how that other kind of forming the principles of justice can be more effective.

The ninth chapter is: "The original position and Scanlon's contractualism" written by Joshua Cohen. His aim is to present Rawls's Rational Advantage Model (his name for the concept) that is used in the original position while analysing the original position as a device of representation. He also compares Rawls's approach with Scanlon's approach through the usage of judgemental (Rawls) and substantive (Scanlon) individualism. Cohen starts by examining how the original position serves as a device of representation. He does this by analysing how usage of the veil of ignorance constrains our rational choice in the original position and how that rational choice promotes one's advantage is a society of free and equal persons who cooperate. The question is why use the Rational Advantage Model and the veil of ignorance. That is why Cohen proceeds to give a more detailed examination of the usage of the veil of ignorance. He tries to show that Rawls doesn't give sufficient reason for employing the thick veil of ignorance, and also tries to show that the ethical interpretation of the initial situation preserves the judgemental individualism in a contract view. He puts substantive individualism in contrast with Rawls's view, stating that his view satisfies only one version of substantive individualism, and that Scanlon's version (the Reasonable Objection view) raises a problem for the Rational Advantage Model. Cohen moves on to another objection of the Rational Advantage Model, which argues that the usage of the original position makes us use additional steps in the argumentation because we need to put in additional facts to the process (Cohen presents them as Further Facts). This raises the cost of using the original position. Cohen illustrates this by showing that the argument from self-respect can be deduced more directly without the use of the original position. Cohen finishes with the analysis of Scanlon's Reasonable Complaint Model and compares it with

the Public Reason Model, which is a part of Rawls's ethical interpretation. He argues that advantages that we can get by using both the Reasonable Complaint Model and the Rational Advantage Model are too costly or illusory, while arguing that the Public Reason Model represents normative political arguments in the best way. Cohen manages to present how the use of rational choice model in the device of representation carries problems that can be seen by closely examining how it works. The biggest problem is that it does not allow us to directly use our reasons for introducing our complaints or arguing for more equality in our society. He also shows that Scanlon's model doesn't solve those problems, because it mostly shifts focus from the rational to the reasonable while blocking some other useful features of Rawls's model.

In the tenth paper Andrews Reath explores "The 'Kantian roots' of the original position". Reath analyses the links between Kant's moral conception and Rawls's theory, while analysing similar concepts and the inspiration Rawls found in Kant's work. He starts by analysing Kant's concept of moral autonomy that is necessary to show how it forms the root of the original position. Reath gives an overview of the Formula of Universal Law, the Formula of Humanity, and shows how they pair up with the principle of autonomy in order to make the categorical imperative necessary and publicly acceptable. That autonomy gives us a part in the making of the universal law and gives us the basis of dignity which enables us to follow the Formula of Humanity. The analysis of Kant's moral constructivism shows that Kant also idealizes persons with specific rational and moral capacities and Reath uses Rawls's analysis of Kant's usage of the rational and the reasonable in order to show connections between their constructivisms. He also examines Kant's conception of a person which is the centre part of his conception, and examines three basic notions of Rawls (well-ordered society, the idea of the person, and the original position). Reath then analyses the Kantian roots of the original position, mostly through the analysis of the person as free and equal, while putting it in comparison with Kant's notion of autonomy. He also points to "parallels between Rawls's original position as a procedure of construction and Kant's CI-procedure" [215] and how justice as fairness contains different forms of autonomy that Kant stipulates. Reath gives an overview of Rawls's first appeal to Kantian interpretations and concepts in the 40th chapter of TJ and how he elaborated the use of Kant's theory in the "Kantian Constructivism". He concludes the paper by pointing to some other parallels between Rawls's theory and Kant, namely the way in which "the reasonable frames the rational", and the similar way in which the original position and Kant's conception lead to substantive normative principles. Reath ends the paper with a possible difference between Rawls's theory and Kant's conception, which arises from different bases on which they build their conceptions of persons. Reath provides us with an insightful analysis of Kantian roots of the original position, while also giving an overview of Rawls's interpretation of Kant's moral theory. This is very useful in examining the setup of the original position and Rawls's motivation for using the constructivist approach in his theory. Reath also shows us how all the concepts that Rawls uses result in the two principles of justice he advocates and how that correlates to Kant's effort of establishing universal moral rules through categorical imperatives.

Paul Weithman is the author of "Stability and the original position from the *Theory* to *Political Liberalism*", the eleventh paper in the collection. He argues that the idea of self-stabilization of the principles of justice and the "self-enforcing argument" can provide answers to important questions concerning Rawls's theory. Weithman first gives a brief view of the original position as a device of representation, because he later argues that it is an integral part of the self-enforcing. After that he lays out the three conditions of self-enforcement on the basis of theory of non-cooperative games (games that do not have a mechanism of coercion), and proceeds to apply them to agreement on Rawls's principles of justice. That leads us to three conditions being modified for Rawls's theory to be self-enforcing agreement: "(R1) The principles of justice that members of the well ordered society are to follow must be specified by terms of an agreement or contract among them...", (R2') None of the members of the well-ordered society can have sufficient reason to deviate from principles of justice, at least so long as all the others comply with them, and all do comply...", and (R3') The fact that the principles would be agreed to in the contract referred to in (R1) must be what brings it about that members of the well-ordered society comply with them, as (R2') requires, and the connection between the hypothetical agreement and the conduct referred to by (R2') must itself be established in ways which treat members of the well-ordered society...". Each condition ends with the same citation from TJ (p. 19) "as moral persons, as creatures having a conception of their good and capable of a sense of justice". Weithman tells us that satisfying conditions (R1), (R2') and (R3') would be, as he calls it, "Rawlsian self-enforcing". He proceeds to argue that the original position satisfies the conditions (R1), (R2') and (R3'). The biggest challenge is to show how the hypothetical agreement could cause compliance, and Weithman does that by arguing in favour of educating citizens through institutions by appealing to the publicity condition and the four-stage sequence. Having satisfied conditions of Rawlsian self-enforcing, justice as fairness becomes stable through self-stabilization (because the citizens themselves maintain justice of the basic institutions). After examining the self-enforcing, Weithman proceeds to present and analyse Rawls's turn to political liberalism. In short, Rawls found inconsistencies in the TJ, and dealt with them by changing the basis of justice as fairness for the moral personality to political personality or citizenship (the shift is from the equal and free persons to citizens). Weitham examines the changes this makes to conditions of Rawlsian self-enforcing by presenting their political variants. By doing that he also explains how Rawls changed his view and some part of the argumentation to accommodate for the changes he made (for example the idea of an overlapping consensus which is an integral for arguing that political liberalism still accommodates the conditions of self-enforcing). Weitham concludes with examining the necessity of the original position. He argues that the original position is essential for providing stability and self-enforcement in the later stages, which makes it crucial for justice as fairness. Weithman manages to shed light on Rawls's shift from TJ to Political Liberalism, while also providing additional arguments that strengthen the stability of the conception of justice Rawls advocates. He manages to do that by closely analysing the concept of self-enforcement, while showing how Rawls's theory is self-enforcing in both forms.

In the last paper Gillian Brock gives an overview of the original position in *The Laws of Peoples* (also the title of the paper). Brock's aim is "high-lighting the role played by the original position in arriving at guidance in international affairs." [247] She starts by briefly introducing the concept of the original position in TJ and then moves on to outline Rawls's Law of Peoples and its eight principles. There are three parts of further examina-tion and they follow Rawls's applications of the original position in order to get to the Law of Peoples. The first one is applied to liberal people in determining their basic structures (internally) and the second one is ap-plied during the process of agreement between representatives of liberal peoples (internationally). At the international level, the parties would come to agree, according to Rawls, on eight principles and three international organizations (concerning fair trade, concerning banking, and one similar to the United Nations). He applies the original position to the agreement between "decent peoples" (peoples that are not liberal but satisfy four condi-tions Rawls specifies for them to be dubbed as decent), while arguing that they will come to the same conclusions as the liberal peoples (also describ-ing a hypothetical decent people Kazanistan). Liberal and decent peoples form a society of well-ordered peoples, based on mutual respect, but Rawls still does not think that his principles of justice should be applied globally. After presenting the shortly described outline, Brock turns to critical re-sponses to the Law of Peoples, and they are: Pogge's objection concerning the international borrowing privilege and international resource privilege that benefit wealthy and powerful states, Rawls's notion of separateness which is opposed by globalization and integration, Rawls's unclear notion of a people, Rawls's exclusion of greater economic inequalities in the world and notion of human rights in the decent states. All this amounts to critics saying that Rawls's theory isn't realistic enough and is not utopian enough, although Rawls claimed that his Law of Peoples makes a realistic utopia. Brock proceeds to defend the Law of Peoples by appealing to Freeman's re-sponse that the Law of Peoples tries to shape foreign policies of the liberal peoples. Decent peoples serve as a theoretical construct that enables us to see with which peoples should liberal peoples cooperate with, and how to establish peaceful and stable world order. She proceeds to present counter-arguments, while claiming that Rawls endorses an incomplete list of human rights. Brock ends with arguments against claims that Rawls has an inad-equate approach to addressing global poverty and that he tolerates non-liberal societies in a problematic way. First counter argument is a fairly simple one: Rawls thinks that giving resources to the poor is not enough. We should aim to reform their institutions in order to give them opportunities of effectively using resources. The second counter argument, in short, tells us that there might be other conceivable ways of developing, ways that don't include liberal society. Brock manages to tackle serious objection to the Law of Peoples by turning our attention to a more realistic goal of peace and stability, the goal that is also a condition for global justice. She manages to defend Rawls's claims about global poverty and human rights by showing us that he gives us more solutions that it may seem at first.

The Original Position represents a valuable contribution to critical ex-amination of a theory of justice put forward by John Rawls. It gives us a clear and simple overview of Rawls's concepts by examining their sustain-

ability, through different examinations of critiques, roots and different application of Rawls's central concept, the concept of the original position. The book is suitable for people (in my opinion, especially students) who want to learn more about the foundations of the contemporary political philosophy but also for advanced readers and professors who want to tackle serious problems that are discussed in the collection. Hinton managed to collect papers that give a great overview because they cover a multitude of approaches to the original position.

<div align="right">

DORIJAN ŽUNIĆ
University of Rijeka, Rijeka, Croatia

</div>

Amy Kind and Peter Kung (eds.), *Knowledge Through Imagination*, Oxford: Oxford University Press, 2016, 251 pp.

Imagination has become a fashionable topic, and its role in procuring knowledge has become a central challenge in the analytical debate on imagination (see, for instance, the 2006 issue of *Metaphilosophy* under the same title as the present collection, *Knowledge through imagination*). The present collection offers a well-organized range of interesting and challenging contributions. They are divided into three groups, the first encompassing taxonomical and architectural issues (featuring papers by M. Balcerak Jackson, P. Langland-Hassan and N. Van Leeuwen), and the second offering "optimistic approaches" (T. Williamson, J. Jenkins Ichikawa, the co-editor A. Kind herself, and J. Church). The optimism is balanced in the third part, featuring "skeptical approaches" by H. Maibom, Sh. Spaulding and by the co-editor P. Kung. I shall choose a paper or two from each group, with apologies to the rest of the authors. (For quotations, I put page numbers in brackets.)

Let me start with the "Introduction" by the editors. They note that "the puzzle of imaginative use concerns two distinct and seemingly incompatible uses to which imagination is often put. (1) Sometimes it is an escape *from* reality, and sometimes it is "used to enable us to learn about the world as it is, as when we plan or make decisions or make predictions about the future. But how can the same mental activity that allows us to fly completely free of reality also teach us something about it?" (Ibid.) How is the "instructive use" of imagination possible? The editors optimistically hope that a closer analysis will explain the joint possibility of the two uses, in particular the instructive one, and see the key to the explanation in constraints that thinkers-imaginers put upon their activity. The constraints come in two kinds. First, they "may be architectural; that is, they may result from our cognitive psychological architecture." (22) Second, the constraints may derive from more spontaneous sources, such as limitations that we voluntarily impose upon our imaginative projects (Ibid).

Amy Kind develops these ideas further in her paper "Imagining Under Constraints". She offers a characterization of imagining that involves a more active effort of mind than does supposition or entertaining a proposition (148), and quotes Kendall L. Walton's (1990) classic *Mimesis as Make-Believe: On the Foundations of the Representational Arts* (Harvard

University Press), suggesting that imagining "is doing something with a proposition one has in mind", Walton, p. 20, (Ibid.). She then proposes a conception of "ideal imagination" modeled on an entertaining science fictional story in which highly developed computing machines predict things in a cold, perfectly calculated way, marching step by step, with "irresistible steps". They obey the "reality constraint" in representing things, and the "change constraint": "when their imaginative projects do require them to imagine a change to the world as they believe it to be, they are guided by the logical consequences of that change (151). She then mentions Tesla and Temple Grandin as human quasi-ideal imaginers. Her conclusion is optimistic: "in modeling our imagination on the ideal imagination of the machines, we are able to make epistemic progress the way they do, by steady, irresistible steps" (159).

Other authors on the optimistic side take similar steps, specifying the constraints imposed upon imagination. Peter Langland-Hassan in his rich paper "On Choosing What to Imagine", concentrates on imaginings that are voluntarily and suitable for guiding action and inference. He lists three essential components that guarantee the guiding power, first, the availability of (top-down) intentions to start imagining, second, of lateral constraints that govern the development of the imagining, and third, the possibility of cyclical interventions by the subject and her intentions, in particular during a given imaginative episode (81).

In his contribution "Knowing by Imagining" Williamson joins the optimistic crew and proposes a cognitive view of imagination, without forgetting its practical value i.e. the importance of practical matters (124); he talks about "a wide range of possible ends"   and a possible practical evolutionary origin of imagination. Also, in his view fiction is not central for imagination, as he pointedly remarks in the concluding sentence of his paper: "… if we try to understand the imagination while taking for granted that fiction is its central or typical business, we go as badly wrong as we would if we tried to understand arms and legs while taking for granted that dancing is their central or typical business"(131).

Among cognitive functions the prominent ones are raising possibilities and assessing the truth-values of propositions (115). This requires cognitive qualities, like rational responsiveness to evidence (116) and the capacity to develop adequate scenarios: the imagination develops the scenario in a reality-oriented way, by default (116). Williamson does not call them epistemic virtues, but this is how a friend of virtue epistemology would describe them. They offer reliability: "…under suitable conditions, the method constitutes a reliable way of forming a true belief as to what would happen in hypothetical circumstances" (117).

Williamson wisely stresses similarities between various exercises of imagination, using them to suggest that the most sophisticated among them, like thought experiments, are nothing special and mysterious. What about science? Williamson has a fine optimistic argument in favor of the serious epistemic status of imagination in it: "One might suppose that, as science progresses, the role of the imagination will increasingly be confined to the context of discovery, and that in the context of justification it will gradually be replaced by more rigorous methods. But there is evidence to the

contrary. For rigorous science relies on mathematics, and so indirectly on the axioms or first principles of mathematics. But when one examines the justifications mathematicians give of their first principles, such as axioms of set theory, one finds unashamed appeals to the imagination"(123). He also stresses that thought experiments are part and parcel of the normal functioning of imagination: "We simply reserve the term 'thought experiment' for the more elaborate and eye-catching members of the kind." So much for Williamson's cognitive view of imagination in general.

The first issue that arises for the project is the classical philosophical one: what is imagination and what is the role of image in it? How close is it to belief? The term "cognitive" seems to suggest a very high degree of closeness; what about the differences? Take imagining a golden mountain: many people will stress the image in such an imagining, but how important is it exactly? Williamson notes that many of his examples "appear to involve an essential role for mental imagery, in some sense" (118), but he quickly adds that "… we should not over-generalize to the conclusion that all imagining involves imagery" (Ibid.). And in fact, he present the imaginative exercise differently, more as a matter of logic and even almost exclusively as a matter of logic and possibly quite sophisticated and complicated, with the full range of tableau methods in the foreground, continuing the venerable tradition of Jaako Hintikka interpreting Kant's notion of *Anschauung* (in his 1969, "*On Kant's Notion of Intuition* (Anschauung)", in T. Penelhum and J. J. MacIntosh (eds.), *The First Critique*, Wadsworth Publishing).

On the other hand, here is how in his central example he presents the way people imagine. He invites us to think of a hunter who finds his way obstructed by a mountain stream rushing between the rocks (117). The hunter "imagines himself trying to jump the stream" (119) and presumably asks himself *If I try here, what is it going to be like?* Williamson notes that "he also has to look carefully at its banks in front of him, to tailor his imaginative exercise as exactly as he can to their actual contours" (Ibid.). But this tailoring of one's imaginative exercise to the contours perceived sounds a bit like creating a visual-kinesthetic moving picture, a video: *it will be like this*. (This is what is often called a *mental model* of the situation, and here imagistic, video-like properties might help a lot.) So, even if we accept that image-producing is not a necessary feature of imagination, it could be a centrally important one, and the non-image-involving cases might be a bit marginal. In general, judgments are easy to elicit with concrete examples. With naive subjects it is the *only* way. However, Williamson stresses the importance of deductive logic and the "tendency of imagination to use something like rules of deductive logic…" (123). He notes "the role of the imagination as a standard means for evaluating conditionals and modal claims (Ibid.) This raises the important issue of the role of logic in relation to imagistic cognition. Like Peter Langland-Hassan**,** Williamson wants to combine the two, and it will be interesting to see what the results in his subsequent work will be. So much about the optimists.

On the skeptical side, the most direct challenge to the project of finding constraints that would rehabilitate imagination is to be found in the paper by Shannon Spaulding: "Imagination Through Knowledge". On her view, the puzzle of how we arrive to knowledge through imagination suggests that imagination is "not sufficient for new knowledge" (222). The argument

seems to be the following: if imagination is to be constrained by extra-imaginative pieces of information and by other abilities, then imagination does not bring new knowledge. But this is too severe a demand. Compare physical constraints. I commute from my home town to my working place about a hundred miles away. For the car to bring me to work there should be a well-established and well-kept road, constraining the travel, there should be red lights helping to prevent crashes, and so on. Imagine someone arguing that therefore "the car is not sufficient" for commuting, and is not doing any real work! Well, the fact that an item needs constraints to function properly does not entail that it never performs any function.

Spaulding has an auxiliary argument: "I have argued that the cognitive capacity to imagine scenarios is distinct from the cognitive capacities that underlie our ability to judge the accuracy of our imaginings" (222) and "… there is nothing in the capacity of imagination itself that could evaluate the accuracy of the possibilities we imagine." (Ibid.) Indeed, there is nothing in the car itself that recognizes the red/green light. This does not show that the car will not take me from home to work, only that car *alone* will not do the work. So much about Spaulding's direct challenge to the instructive use of imagination.

Let me mention, however, that in her text the challenge is preceded by a rich and very provocative analysis of one particular kind of imaginational enactment, namely simulation. Her argument resembles the general one we just summarized. Her example is the following: I watch John tease Mary, and try to figure out why he is doing this. I simulate his activity, and end up concluding that John likes Mary and is trying to get her attention. Fine, but how do I choose this option rather than some other, equally plausible in itself, for instance that he is just humiliating her? I need additional information, and my simulation tells me nothing about these matters. Again, to me it looks like simulation has done the main job, like the car in our example; the fact that the main job cannot be fully accomplished by the main agency in question tells little against it.

Heidi Maibom's paper "Knowing Me, Knowing You: Failure to Forecast and the Empathic Imagination" joins in with bad news about people's abilities to recognize their own characteristics and attitudes, and abilities to project items of self-knowledge onto their neighbours.

Peter Kung's "Thought Experiments in Ethics" is not so generally pessimistic as the papers by the two preceding authors. He just warns us that typical ethical thought experiments, especially ones that are meant to produce counterexamples to crucial ethical claims, CTEs for short, are organized around sharp, binary divisions, offering "forced choices with fixed outcomes": would you pull the lever, and kill three people, but save five, or not? He develops his criticism in a rich and subtle way, connecting it with issues of imagistic (he calls it "pictorial") vs. non-imagistic representations, with topics of modality and so on. He claims that "imagining CTEs gives us *no reason* to believe that forced choices with fixed outcomes are genuine possibilities" (228, italics mine). We should use more realistic scenarios in our thought-experiments.

Let me note that real life often does offer "forced choices with fixed outcomes": would you marry the person you are so passionately attracted to,

but whom you realize to be a very dangerous partner, or not?, would you vote for Trump, for Clinton or for Sanders, or not vote at all? So ethicists might hope to offer some answers to people facing such choices, and they might prepare themselves by going through imaginary exercises featuring them.

Let me conclude that the optimistic side might have chances to survive. And let me add the following: if we accept that imagination follows real-world (or quasi-real-world) constraints, the question arises where the representations of the constraints come from. One possible unitary answer is that thinkers have mental models of reality, and that, when they ask themselves an instruction-oriented question, the models available to them constrain their subsequent imagination. If the result is worth remembering and taking into account, it can be integrated back into one of the models, so that in the future it will provide a relevant "lateral constraint" to some exercise of imagination. If we assume that imagination is typically imagistic, and that mental models are typically concrete and "iconic", but that both allow for thought processes that range from more iconic-pictorial to more digital deductive ones, then we shall notice that the two media, imaginational and model-sustaining one, nicely fit together and can interact in a non-problematic way.

NENAD MIŠČEVIĆ
University of Maribor, Maribor, Slovenia
CEU, Budapest, Hungary

Instructions for Contributors

All submissions should be sent to the e-mail: cjp@ifzg.hr. Submissions must be in English and formatted to be double-spaced with suitably wide margins, an A4 page size, and automatic page numbering.

Articles are normally no longer than 8,000 words of main text including bibliography. The Journal will consider longer papers, but, once these limits are exceeded, authors should bear in mind the editorial policy that the acceptance bar raises with increasing length.

Manuscripts should be compiled in the following order: cover page; title; abstract (not exceeding 200 words); keywords (3 to 6); main text; appendices (as appropriate); references.

All the authors of a paper should include their full names, affiliations, postal addresses, telephone and fax numbers and email addresses on the cover page of the manuscript. If a paper is co-written, one author should be identified as the Corresponding Author. The cover page must be submitted as a separate document. All submitted manuscripts must be prepared for blind review, with revealing acknowledgements and self-identifying references removed.

Sources are cited in the text by the author's last name, the publication date of the work cited, and a page number if needed, e.g. (Barber 2007: 324). Full details appear in the reference list in which the year of publication appears immediately after the author's name:

Barber, A. 2007. "Linguistic Structure and the Brain." *Croatian Journal of Philosophy* 21 (7): 317–341.

Williamson, T. 2013. *Identity and Discrimination*. Oxford: Wiley-Blackwell.