Slurs, Truth Conditions and Semantic Internalism

Michael McCourt
Visiting Assistant Professor, University Honors Program
The George Washington University

Christopher A. Vogel
Independent Researcher

McCourt, M., & Vogel, C. A. (2025). Slurs, Truth Conditions and Semantic Internalism. *Grazer Philosophische Studien*, 101(4), 486-528.
(please cite original source: https://doi.org/10.1163/18756735-00000233)

ABSTRACT

One class of views identifies the derogatory capacity of a slur with its truth-conditional contribution, while a second class denies this. This second class of *pragmatic* views is purportedly burdened with the *non-pejorative counterpart* problem. Pragmatic approaches to slur derogation mandate that each slur has an alternative coextensive non-derogatory expression—a non-pejorative counterpart. We offer a solution to this problem that identifies its source with the underlying assumption that expressions have extensions, an assumption largely independent of any commitments central to pragmatic approaches to slur derogation. If meanings are instructions to build concepts, then neither slurs nor their purported counterparts have extensions, and thereby cannot be coextensive. By treating meanings as instructions, not only can pragmatic views avoid the problem counterparts pose, but such a semantics offers a flexibility that can accommodate multiple pragmatic mechanisms to explain the complex behavior of slurs and their use.

Keywords: Slurs, Semantics, Pragmatics

Slurs denigrate. The use of slurs to denigrate individuals trades on the relationship between the slur and the group it historically targets for derision. Semantic accounts of this feature of slurs identify the derogatory character of slurs with their truth-conditional contribution. In contrast pragmatic accounts deny this truth-conditional claim, typically identifying the derogatory character of slurs with some aspect of their use.

We argue for two claims regarding accounts of the derogatory character of slurs. First, we argue that the *non-pejorative counterpart* (NPC) problem presents a dilemma in accounting for the derogatory character of slurs. While the diverse constellation of semanto-pragmatic properties of slurs generally seem to suggest a pragmatic account of slur derogation (and offense), these views are saddled with the NPC problem. Roughly, pragmatic approaches to the derogatory character of slurs (seemingly) require that each slur has a coextensive non-pejorative counterpart. This requirement leads to implausible predictions about slur behavior. Second, we

contend that a plausible resolution to this dilemma is to deny the general externalist assumption that grounds it. By denying that meanings determine extensions, a pragmatic view of derogation overcomes the NPC problem. Further, taking this internalist option comes at little cost for a pragmatic approach to slurs, since much of the utility of taking meanings to be truth-conditional can be preserved by adopting an analogous view about the concepts that word meanings recruit.

We begin by outlining the variety of properties that characterize slurring expressions, briefly illustrating how these properties differentially suggest semantic and pragmatic accounts, before situating our own proposal within a pragmatic framework (Section 1). We then detail the NPC problem, assess attempts to resolve it, and ultimately show that the NPC problem plagues any pragmatic account which adopts an externalist semantics, underscoring the dilemma the NPC problem poses (Section 2). We then articulate and defend a pragmatic approach to derogation that denies the externalist claim that the meanings of natural language expressions determine truth-conditions. We sketch an internalist framework that identifies meanings as instructions to retrieve and build concepts (Section 3) and illustrate how this basic proposal is compatible with a range of views about the derogatory character of slurs (Section 4). This internalist approach to slur derogation both resolves the NPC problem and preserves the merits of pragmatic accounts that traditionally assume an externalist semantics. We then discuss how our approach differs from others (Section 5) and respond to objections (Section 6) before concluding (Section 7).

1 Background

A set of observations concerning slurring expressions calls out for special treatment within an overall linguistic theory.[1]

Autonomy:[2] A slur apparently retains its derogatory character[3] even when the speaker does not intend to derogate.

Non-Embedding: A slur apparently retains its derogatory character even when embedded under various syntactic frames: in indirect speech reports, under negation, in conditionals, etc.

Perspective: Use of a slur compels the audience to attribute to the speaker an attitude commensurate with the slur's derogatory character.

---

[1] Similar lists are given in, e.g., Potts 2007 and Bolinger 2017. We assume only that this list is useful in establishing the target for explanation, without assuming that the class of slurs can be defined in a way that separates it cleanly from non-slurring pejoratives like 'jerk.' For discussion of whether and how the category of slurs might be defined separately from closely related notions, see Croom 2011; Anderson & Lepore 2013; Jeshion 2013; Popa-Wyatt 2016; Neufeld 2019.

[2] The label *Derogatory Autonomy* is due to Hom (2008). Anderson & Lepore (2013) call this feature *Derogatory Independence*. Also compare *Offensive Independence* in Bolinger 2017. For discussion of the relationship between derogation and offense, see Bolinger 2017 and Davis & McCready 2020. Our focus in what follows is derogation. In Potts' (2007) feature list—which, in contrast with Bolinger's, is meant to characterize expressive language more generally (to include both slurs and non-slurring pejoratives like 'jerk')—*Independence* refers to the relationship between the expressive and descriptive dimensions of meaning in a bifurcated semantics, a notion of independence which should not be confused with what we have labelled *Autonomy* in this paper. We have chosen to label the feature of slurs, whereby they tend to derogate even when their users intend no such thing, as *Derogatory Autonomy*, a specific version of the more general phenomenon that we might label as *Expressive Autonomy*. (However, we concede that, with a clarifying footnote in relation to Potts 2007, *Derogatory Independence* would do just as well.)

[3] Functions from contexts to contents are often called 'characters.' This specialized use of the term, due to Kaplan (1989), should not be confused with the much more general sense utilized here and throughout the paper. The use of 'character' in what follows pertains to that aspect of slurs that makes them particularly derogatory.

Variation: There is substantial variation from one slur to the next in terms of degree of derogatoriness.
Insulation: In some contexts of use, the derogatory character of a slur can be mitigated or eliminated.
Specificity: The pejorative force of a slur is related to (perceived) membership in a particular social group.[4]

A complete account of slurs and their uses must consistently explain these six observations. However, these observations push in inconsistent explanatory directions.

Accounts of the derogatory character of slurs fall into two groups. One type of explanation identifies the derogatory character of a slur with some part of its meaning. The alternative approach locates derogation elsewhere—not as part of slur meanings, but as a feature of their uses. In what follows we call views in the first class 'semantic' and views in the second class 'pragmatic.' Our use of these terms is restricted to distinguishing two different ways of explaining the derogatory character associated with slurs: semantic explanations appeal to the truth-conditional contributions of derogatory word meanings, while pragmatic explanations do not. For example, a semantic explanation might take a slur like 'mick' to be truth-conditionally equivalent to 'Irish and therefore contemptible.'[5,6] Alternatively a pragmatic explanation might claim that a *speaker* who chooses 'mick' in place of 'Irish' signals a derogatory attitude, even though the meaning of 'mick' lacks any derogatory content.[7]

---

[4] Jeshion's (2013) labelling and discussion of G-referencing, G-extending, and G-contracting uses of slurs clearly illustrates, by way of the logical constant, G, the feature we label as Specificity. (For a brief illustration of Jeshion's categories, see pp. 3-4 of this manuscript.) Many lists (including those in Potts 2007 and Bolinger 2017) take Specificity for granted. However, we take Specificity to be crucial enough to list explicitly, since it poses a substantial problem for views, like our own, that divorce linguistic meaning from reference and satisfaction. See more discussion on this point in Section 6 below.

[5] In most of what follows, we use the somewhat archaic and moderately derogatory ethnic slur, 'mick,' in our examples. The reason for this choice is related to Autonomy and Variation. To report our own phenomenology, even in the clearly pedagogical context of an academic paper, reading (and typing) an especially forceful slur may warrant some degree of offense. Since our argument does not trade on any feature that is unique to particular slurs, little seems lost by choosing an example that offers less moral risk, given its archaic status. However, what might be lost on some English speakers is how the slur 'mick' is used. The etymology of the term in American, Canadian, and British English as pertaining to the Irish stems from a diminution of 'Michael,' a common Irish name in the 19th century when the term first emerged. It is speculated that this usage may have been reinforced by a common feature of Irish surnames, which often enough begin with 'Mc' and 'Mac.' (Also note that, in Australian slang, 'mick' has been used not as an ethnic slur but instead as a slur for Catholics of whatever ethnicity.) See also the early 20th century phrase 'slip someone a mickey,' which describes an act of surreptitiously adulterating a beverage with a drug (McCabe 2008) and is parasitic on an earlier (chiefly Canadian) use of 'Michael' and later 'mickey' to refer to a small bottle of alcohol.

[6] Hom (2008) articulates such a naïve version of a semantic approach, but only to replace it with a more sophisticated version, discussed below.

[7] See also Bolinger 2017. – As we organize the space of views, all and only accounts on which a slur itself encodes derogatory content are semantic. This means, for example, that we would count as pragmatic a "radical contextualist" approach on which a slur, much like a standard indexical, encodes not a derogatory content, but a function from contexts of utterance to a (possibly derogatory) content (Kennedy 2002). Given the many purposes they have served in the past, any use of 'semantic' and 'pragmatic' is apt to generate some confusion. However, we limit their use in what follows to the task of distinguishing two types of explanations for the derogatory character of slurs: those that find derogation in encoded word meaning, and those that do not. The view we defend here is pragmatic. For reasons of space, it is not possible to discuss here all of the accounts of slurs that are semantic, pragmatic, or hard to categorize; and, given that our present aim is only to articulate a novel pragmatic account of derogation, it is also unnecessary for us to offer any such exhaustive overview. For our purposes, it is enough to

Some of the observations above appear to favor semantic approaches; others recommend pragmatic approaches. Take, for example, Autonomy. If one uses a slur, that use will generally involve derogation of the target group regardless of the speaker's intent.

(1) # They are micks, but I've got nothing against the Irish.

The infelicity of the sentence in (1)—marked by the octothorpe—illustrates the autonomy of the slur 'mick,' which retains its derogatory character despite the explicit disavowal in the conjunctive continuation. Semantic accounts straightforwardly explain Autonomy, on the plausible assumption that speaker intentions do not *ipso facto* dictate the content of an expression, even on an occasion of use. According to a semantic approach, (1) determines (relative to an utterance context that assigns referents for 'they' and 'I') a derogatory proposition. On a pragmatic approach, Autonomy cannot be explained as straightforwardly. A use of (1) would determine no derogatory content in the absence of those pragmatic features relevant to a slur's ability to derogate—features over which the speaker *does* have considerable authority.

Additional derogatory uses underscore the Autonomy of slur derogation, as illustrated in (2) and (3).

(2) He isn't Irish, but he's still a mick.

(3) He is Irish, but he's not a mick.

Jeshion (2013) contrasts such "G-extending" and "G-contracting" uses in (2) and (3) respectively with the kind of "G-referencing" use that is illustrated by (1)'s first conjunct.[8] Exactly how to understand (2) and (3) is subject to dispute. While some theorists argue that such uses are non-literal (Anderson & Lepore 2013, Jeshion 2013), others have cast doubt on this strategy and argued that a theory of slur derogation should be able to accommodate the intuition that sentences like (2) and (3) can be felicitously and non-figuratively used to convey derogatory messages (Cepollaro 2017, Neufeld 2019). G-referencing, -extending, and -contracting uses all involve derogation in a way that highlights Autonomy: a speaker of (1-3) will be taken to express a derogatory attitude toward Irish people, even if the speaker (explicitly) lacks such an attitude. At least initially, this observation seems to favor a semantic account on which derogatory content is somehow attached to the slurring expression itself.

Insulation pushes in the other direction. Not all uses of slurs involve derogation.[9] Among non-derogatory uses of a slur, we can distinguish several types. If one Irishman utters just the first conjunct of (1) to make a claim about some fellow Irishmen, signaling affection rather than contempt, this would be a case of *appropriated* or *reclaimed use*.[10] Other uses indicate that a

---

establish the basic divide and to locate our proposal on the pragmatic side of it. In Section 5 we discuss how are view differs from some closely related proposals on each side of the semantics/pragmatics divide.

[8] Compare Croom's (2015a) discussion of non-target uses of slurs.

[9] Since we focus here on *uses* of slurs, we set aside consideration of mere *mentions* (for example, in dictionaries or in the didactical context of an academic paper about slurs). While mentions may generate and perhaps even warrant offense, a mere mention considered by itself is clearly non-derogatory. We thus set aside all such mentions of slurring terms, focusing on cases of insulated use.

[10] Due to limited space, important questions concerning insulation cannot be adequately addressed in what follows. For example, we mostly set aside debates about whether there is a single mechanism for reclamation, or several (though see Section 4.1 for a brief discussion of how our CAP proposal relates to such debates). We also mostly set aside discussion of subvarieties of reclaimed use. In some cases, reclamation shifts the attitude associated with reference to the target group from negative to positive in order to pursue broader social change, as (arguably) occurred with the use of 'queer' (Brontsema 2004). In other cases, the shift to a positively valenced attitude in association with a slurring expression involves no ambition for broader social change, as in certain uses of the n-

positive attitude is not required for an in-group speaker to avoid derogation. Drawing a comparison to affectively neutral uses of 'buddy' and 'man,' Anderson (2018, 9) labels these *referential uses*. Yet another subclass of non-derogatory in-group uses is distinguished from the foregoing in Zeman (2022), which centers around the Romanian slur 'țigan' (roughly translated as 'gypsy'). According to Zeman, 'țigan' permits of a non-derogatory use by members of the Roma community that the slur historically targets. However, in contrast with Anderson's (2018) referential uses, which involve reference to someone other than the speaker, non-derogatory uses of 'țigan' are often self-referential, or *identificatory uses* (Zeman 2022, 949).

In other contexts a slur can be used without derogation in an apparently non-appropriative way, even by out-group members, as illustrated by (4).

(4) Institutions that treated Irish immigrants as micks were morally depraved.

Adapted from Hom (2008), who labels these *non-derogatory, non-appropriated* (NDNA) uses, an utterance of (4) in a pedagogical situation seems to lack derogation. Further, out-group *metaphorical use*s seem to avoid derogation, as illustrated by the John Lennon and Yoko Ono (1972) single, cited by Zeman (2022, 945), in which being a woman is figuratively identified with being African American via an apparently metaphorical use of the n-word.

This brief overview of (plausibly) insulating, non-derogatory uses of slurs seems to favor pragmatic accounts that locate derogation not in the meanings of slurs, but rather in facts about their use. At first glance, a single slur can carry derogatory force or not, can express a negative, positive, or neutral attitude, and is amenable to non-derogatory, outgroup member deployment or not—all depending on features of the situation in which the expression is used. The purpose of raising these considerations is not to argue that they decisively favor pragmatic approaches. Our goal here is not to argue that a pragmatic approach to explaining the derogatory character of slurs is obviously the better option in light of available observations and arguments. Responses and counterarguments can and have been offered on behalf of the semantic approach to the challenges raised here. For example, proponents of the semantic strategy have appealed to ambiguity (e.g., Jeshion 2013, Sennett 2016) to explain the data. If some allegedly non-derogatory uses of a slur in fact involve a different, non-slurring term, this reduces the size of the challenge faced by semantic accounts. In addition, any pragmatic account will face challenges of its own, especially in relation to the observed Autonomy of slurring expressions, whereby derogation often emerges even in absence of derogatory intent on the part of the speaker.

We have sketched the broader dialectic simply to locate our proposal within it. Resolving this conflict between the observations from Autonomy and Insulation (and other aspects of slur usage) presents a central problem in offering an account of slurs. The main challenge for proponents of the semantic approach, on which derogation is part of a slur's truth-conditional contribution, is to explain Insulation (and Variation) without abandoning their core commitment that a slur itself encodes derogatory content. The central challenge for a proponent of the pragmatic approach is to tell a plausible pragmatic story (for example, in terms of implicature or presupposition) that explains the observations of Autonomy, Non-Embedding, and Perspective. Our aim in this paper is to offer a novel way for proponents of the pragmatic approach to meet this second challenge.

Beyond the need to offer a consistent explanation for these observations, pragmatic accounts apparently face a direct challenge that gives rise to a central dilemma in accounting for slur derogation—what we call the *problem of non-pejorative counterparts*, or *NPC problem* (see

word (e.g., Kennedy 2002). Jeshion (2020) labels these two subvarieties as *pride* and *insular reclamation* (see also Bianchi 2014; Anderson 2018). See also a brief but helpful overview in Zeman 2022: 947-949.

also Hom 2008, Williamson 2009, Hom & May 2013). Pragmatic accounts seemingly require that every slur has a *non-pejorative counterpart*, an expression that is coextensive with the slur in question but that lacks the pejorative baggage. For example, 'mick' might have as its NPC the neutral term 'Irishman.' This coextensive commitment, according to the NPC problem, makes implausible predictions about the truth of sentences containing slurs (and their NPCs). For example, if 'mick' and 'Irishman' are coextensive, then 'Every Irishman is a mick' is trivially true.

2 Pragmatic approaches and the NPC Problem

Any account that denies that a slur itself encodes derogatory content is pragmatic and, as we illustrate in this section, must resolve the dilemma posed by the NPC problem. Our goal in this section is to indicate the space of views adequately enough to establish the basis of the NPC problem. Given this aim, the overview to follow is not exhaustive. Rather, the purpose is to describe several pragmatic approaches to slurs and derogation, highlighting the features they share that are pertinent to generating the NPC problem.[11]

One subclass of pragmatic accounts appeals to a difference in tone between a slur and a neutral (non-derogatory) term for the same group (Dummett 1981; see also Frege 1892, Lepore & Stone 2018). Another (large) group of pragmatic accounts explains derogatory character in terms of Gricean notions of implicature (Grice 1975). Some of these views appeal to something like conversational implicature, leveraging the speaker's use of a slur, as opposed to a non-slur, as evidence of the proposition or attitude the speaker intends to communicate (Bolinger 2017, Camp 2013). Others appeal instead to the notion of a conventional implicature, whereby in using a slur a speaker communicates, without asserting, a derogatory message that cannot be cancelled, as illustrated by the initial example (1), repeated here (McCready 2010; Whiting 2008, 2013; Williamson 2003, 2009, 2010).

(1) # They are micks, but I've got nothing against the Irish.

Yet another subclass of pragmatic accounts attributes a slur's derogatory character to a presupposition, whereby in using a slur the speaker takes for granted various derogatory beliefs about the group targeted for reference (Cepollaro 2015). Lastly, a wide range of expressivist accounts appeal to pragmatic mechanisms such as presupposition (e.g., Lasersohn 2007, Sauerland 2007, Schlenker 2007) or conventional implicature (e.g., Potts 2007) to explain the non-truth-conditional message speakers express by using a slur.

These pragmatic accounts share the core commitment to derogation being a feature, not of a slur *per se*, but rather of its uses, and so face the challenge of balancing the explanatory demands of the relevant observations without giving up that commitment. If Autonomy is removed from the list of explananda, the pragmatic approach seems clearly favored over a semantic approach that locates derogation as part of the meaning of a slur. The central challenge for pragmatic approaches is to somehow account for Autonomy without giving up on a general pragmatic approach to explaining derogation. This is unsurprising given that the aim of pragmatic approaches is to extricate the derogatory character of slurs from their meanings, identifying slur usage as the culprit.

A pragmatic account of derogation, by extricating any derogatory content from word meaning, seems committed to the following claim:

[11] For an indication of the full space of pragmatic theories of derogation and the differences among them, see Finkbeiner et al. (2016).

(NPC) For any slur σ, there is a non-slurring expression η such that σ and η are coextensive.

The argument that purportedly saddles pragmatic accounts with accepting (NPC) is straightforward. Pragmatic accounts identify the derogatory character of a slur with aspects of the slur's use, denying that slur meanings contain derogatory content. Granting that slurs function by identifying an extension—the group targeted by slur usage—the meaning of a slur determines an extension in a manner independent of the derogatory character of the slur. As such, there is (in principle) some expression, complex or otherwise, whose meaning determines the same extension as a given slur (see, e.g., Hom 2008, Williamson 2009, Hom & May 2013).

To see a few examples motivating this argument, consider first the Fregean approach, on which (for example) 'mick' and 'Irishman' share a sense but differ in tone (compare 'horse' and 'nag'). Given that the sense of an expression determines its extension, the single sense these words share determines the same extension. The expressions are coextensive, but differ in terms of derogation. Likewise, an implicature account of the derogatory character of slurs seems to be committed to the claim in (NPC). On these views, regardless of the specific pragmatic mechanism that takes an audience from a slur-sentence meaning to the implicated speaker's meaning, the expression alone fails to deliver the derogatory blow. Since accounting for Specificity seemingly obligates any such view to accept that a slur has an extension, in principle some non-slurring expression determines that same extension.[12] Generally, then, pragmatic views seem committed to the co-extensionality of a slur and its neutral counterparts.

The problem with being so committed is that (NPC) makes implausible predictions about sentences with slurring constituents. Consider (5).

(5) All Irishmen are micks.

If 'Irishman' and 'mick' are co-extensional, then (5) is trivially true. However, (5) seems otherwise, in contrast with (6).

(6) All Irishmen are Irishmen.

Perhaps (5) is a covert or "second-class" analyticity (borrowing the notion, but not the application, from Quine 1951, 24). But then it is unclear why every competent speaker of English agrees that (6) expresses a triviality, while many would object to an utterance of (5) as being false (and offensive). Pragmatic views then seem saddled with accepting the implausible consequence that (5) and (6) express the same truth-conditions.

However, proponents of pragmatic approaches deny that such data pose much of a problem for pragmatic accounts. One reply accepts that sentences like (5) and (6) convey identical truth-conditions, and are thus both analytic. Both sentences *say* the same thing in expressing identical truth-conditions, but the difference in speaker intuitions about their triviality is not a matter of *what* is said, but *how* it is said (Whiting 2013). On this approach, what a speaker using (5) says is equivalent to what is said in a use of (6). However, the message that the (bigoted) user of (5) intends to convey differs from what is said, as an implicature generated by

[12] Recent work has targeted the assumption that slurs must have non-pejorative or *neutral* counterparts. Part of the concern regards the notion of neutrality. For example, Ashwell argues that "any term that is co-extensive with the target group for a gendered pejorative … will therefore be nonneutral in two ways: it will be evaluatively nonneutral … and covertly normative" (2016, 238). In terms of motivating the NPC problem, these concerns pose no obstacle, since all that is required is that (at least some) slurs have *coextensive* counterparts that are purportedly reflected in the differing judgments from speakers about truth under substitution. Likewise, whether "the meaning of a slur is identical to the meaning of its neutral counterpart" is orthogonal to whether they are coextensive (Falbo 2021, 3). A pragmatic approach to slurs might well deny this "semantic equivalence thesis," and still be burdened by the NPC problem.

the *way* the proposition is used. For any pragmatic account, a similar story seems available: accept the truth conditional identity of (5) and (6), yet attribute the apparent difference in triviality to the pragmatic feature the view marks as carrying the derogatory character of slurs.

However, such a response seems unmotivated by speaker data. As noted earlier, the derogatory character of a slur seems to scope out of a complement clause position. Reporting what a bigot said using their words does not inure the reporter from disapprobation; it does however *decrease* the perceived offensiveness of the slur (Cepollaro et al. 2019). This data is difficult to explain on the what-vs.-how response to the NPC problem. After all, in cases of reporting *what* another said, the failure to use a truth-conditionally equivalent counterpart should trigger the pragmatic mechanism that marks the distinction between (5) and (6). However, one then wants to know why speakers judge those case as *less* offensive if the *same* pragmatic mechanism—namely, *how* the speaker worded their report—is doing the pejorative work.

But even if we accept the pragmatic what-vs.-how response to cases like (5) and (6) as a general strategy to NPC cases, the response seems implausible. Consider a G-contracting use of a slur, as in (7).

(7) Cillian is an Irishman, but he is not a mick.

If the truth-conditional contribution of 'mick' is identical to its counterpart 'Irishman,' the sentence in (7) should at least seem *somewhat* contradictory. However, not only do such sentences not seem contradictory, it is difficult to decipher how a pragmatic mechanism that masks the purported analyticity in (5) could *also* leverage the (apparently covert) contradiction in (7) to yield a coherent (although bigoted) claim about Cillian's virtues.

The force of the NPC problem is also vivid in NDNA uses, as suggested by the clear contrast between the following:

(8) Institutions that treat persons as micks are morally objectionable.

(9) Institutions that treat persons as Irish are morally objectionable.

A speaker can happily deny the sentence in (9), without thereby accepting the bigoted consequence of denying (8). But if 'Irish' is a truth-conditionally identical non-pejorative counterpart of 'mick', that speaker embraces flatly contradictory claims. This plainly belies speaker judgments, precisely because treating a person as a mick and treating them as Irish are different ways of treating persons. As such the what-vs.-how move is simply not available to explain the difference between (8) and (9). One might think at least some institutions morally *ought* to treat individuals distinctly based on their alignment with historically marginalized groups, yet not be compelled to embrace the bigoted consequences of denying (8).

A further data point that underscores the import of the NPC problem pertains to the propensity, for non-bigots at least, to deny that a slur *ever* aptly applies. This feature is evident when contrasted with other derogatory terms that are not slurs. Compare the following conversations:

(10) A: The factory only employs non-unionized workers.
B: So the factory only employs scabs.
A: # No, no one is a scab.

(11) C: The dairy only employs Irishmen.
D: So the dairy only employs micks.
C: No, no one is a mick.

Despite being a derogatory term for a non-union worker, there seems to be little reluctance to taking 'scab' as applying to people. The same does not hold for the slur 'mick'—a point made more vivid with a less archaic slur. On the pragmatic what-vs.-how reply, speaker C is simply

making a false claim in saying that there are no micks since, on the what-vs.-how reply, there are exactly as many micks as there are Irishmen. That simply misrepresents the data. But even if we accept the claim that there are as many micks as there are Irishmen, one wants to know why we are compelled in the case of slurs to errantly endorse claims to the contrary, while in likewise derogatory non-slur cases such denials are not forthcoming. The same what-vs.-how story, available to explain the misplaced judgment in cases like (11), would apply to cases like (10), despite our willingness to accept the applicability of the derogatory expression.

These observations make for a compelling argument in support of a semantic account of derogation. Semantic accounts attribute the potential for (5) and (6) to differ in truth-value to a difference in their truth conditions. Naturally this difference is explained by offering different satisfaction conditions for 'Irishman' and 'mick,' whereby the latter includes a derogatory component that the former lacks. If we assume that a word's meaning must involve truth-conditional contributions, and thereby determines its extension, then seemingly any pragmatic account we give for the difference between 'mick' and 'Irishman' will be saddled with the troubling consequence that they are truth-conditionally equivalent. Indeed, many proponents of pragmatic accounts explicitly accept the claim in (NPC) (e.g., Bianchi 2014; Blakemore 2015; Cepollaro 2015; Hornsby 2001; Vallée 2014; Williamson 2003, 2009). Accepting the claim in (NPC) thereby burdens pragmatic accounts with explaining the troubling consequence exhibited by examples like (5) and (6).

One possible explanation would appeal to a metalinguistic reading of (5), repeated here.

(5) All Irishmen are micks.

On a metalinguistic reading, (5) would make a true claim not about groups of people, but about uses of words. Compare (12).

(12) No Irishmen are micks, bigots are wrong.

The negation in (12) permits of a metalinguistic reading on which it denies a claim about the applicability or permissibility of the term 'mick.' Arguably all nonpejorative uses of slurs involve such metalinguistic uses, and therefore should not be taken as evidence in favor of a truth-conditional approach to derogation (Anderson & Lepore 2013, 28-29; Jeshion 2013, 259; fn. 47; Bolinger 2017, 459, fn. 14; see also the extensive discussion in Cepollaro & Thommen 2019). Example (5) might likewise be interpreted as a metalinguistic claim, rather than a standard universal generalization, indicating that 'mick' is a term used to refer to the Irish. If (5) makes a meta-linguistic claim, the proponent of the pragmatic approach to derogation can deny that (5) and (6) express the same truth-conditions. However, simply noting that a metalinguistic reading of (5) is available does not avoid the NPC problem. While we are aware of compelling arguments for metalinguistic readings of sentences like (12) (e.g., Cepollaro & Thommen 2019), those arguments do not straightforwardly apply to sentences like (5). The primary reason is that sentences like (5), and the variety of examples that motivate the NPC problem, simply do not seem to be *about* language. But even if meta-linguistic interpretations are available for every problem sentence, that does not show that meta-linguistic interpretations are the *only* available interpretations of them. The cases related to NDNA uses make this point rather forcefully. An Irish person objecting to the government's treatment of them in using (8) above need not be targeting the government's terminology. The moral failing of such institutions is that they treat *people* in unjust ways, often far beyond the use of certain terms to talk *about them*. So even if meta-linguistic considerations might explain away some problematic NPC cases by noting the availability (or even primacy) of meta-linguistic interpretations, as a general diagnosis for the variety of NPC cases above, the strategy seems unpromising.

The NPC problem thus poses a dilemma. On the one hand, the diverse properties of slur derogation arguably outstrip the ability of a truth-conditional account of derogation. On the other hand, any non-truth-conditional view seems to be saddled with the NPC problem, which is far more stubborn than extant replies suggest. So, while semantic approaches can easily avoid the NPC problem (by, say, simply indicating that every slur extension is empty[13]), pragmatic accounts appear to be stuck with it, leaving nothing but bad options on the table. However, it is not the case that pragmatic accounts of derogation, simply in virtue of being pragmatic, must deal with the NPC problem. Notice that an important assumption in the NPC argument is that slurs have meanings that determine extensions. So (for example) even a conventional implicature-based account of derogatory character that accepts the assumptions (a) that word meanings determine extensions and (b) that sentence meanings determine truth conditions will be saddled with the NPC problem, since the conventions that distinguish implicatures from meanings carve out the extension determining features of word meanings. But one way to resolve that problem is to deny the assumptions in (a) and (b). After all, it is possible (a) and (b) are mistaken, and that 'mick' and 'Irishman' fail to share an extension in common because each of them individually fails to have an extension that might be shared. That, in any case, is what we argue, on independent grounds, in the next section.

An internalist approach, whereby word meanings are divorced from extensions, and sentential meanings are divorced from truth conditions, is, as we argue, perfectly compatible with pragmatic approaches to slurs. Further, many of the available pragmatic approaches are not only compatible with, but also naturally recommended by, an internalist semantics. So, the combination of a pragmatic approach to derogatory character with an internalist approach to linguistic meaning both avoids the NPC problem and is independently motivated. Slurs, as natural language expressions, do not have coextensive neutral counterparts because natural language expressions *simpliciter* fail to have extensions. One upshot of denying the assumption that the meanings of slurs do not determine extensions, explored in Section 4, is that this permits a multifaceted pragmatic approach to the behavior of slurs that can capture seemingly conflicting features like Autonomy and Insulation.

3 Semantic internalism combined with a pragmatic approach to slurs

For semantic internalists, the solution to the NPC problem, in brief, is that a given slur has no non-pejorative counterpart. One internalist proposal identifies the meanings of natural language expressions with *instructions* to access and combine concepts. A slur like 'mick' and a purported synonym like 'Irishman' thereby offer different instructions for concept retrieval. So, while these two terms might, on some occasions of use, select concepts with sufficient overlap in content, the linguistic items themselves encode different instructions, and therefore do not share a meaning. Likewise, 'mick' and 'Irishman' cannot share an extension because instructions, quite generally, are not the kind of thing that have extensions in the first place.

Before sketching the internalist framework that underwrites a pragmatic account of slurs, it will be helpful to identify one pervasive phenomenon in natural languages that motivates this internalist view: natural language expressions exhibit remarkable *flexibility*.[14] Consider the noun phrase (13).

(13) the blue book

[13] This strategy is pursued in Neufeld 2019, for example, as discussed below in Section 5.

[14] The observation goes back at least to Aristotle (e.g., *Categories* 8b25-9a28). See also Waismann 1945, Austin 1961, Chomsky 1977, Travis 1989.

A speaker might use the expression in (13) to talk about a variety of worldly objects.

(14) The blue book is in the closet.

(15) The blue book is award-winning.

Used in (14), 'the blue book' regards a *concrete* physical object put into storage, while the use in (15) regards an *abstract* content praised by critics. We can see this flexibility in the various interpretations of the following:

(16) The blue book is Sona's.

On one reading, (16) regards an object produced on a printing press that (say) Sona purchased at a newsstand, while on another reading (16) is about (say) the biography Sona authored. These readings are closely related, exhibited by the fact that a single use of 'the blue book' can be followed up by conjoined predicates compatible with each interpretation (Zwicky & Sadock 1975):

(17) The blue book is in the closet even though it is award-winning.

A speaker of (17) is apparently able to refer with the same noun phrase ('the blue book') to both the acclaimed written content and its physical vehicle. The single use of 'book' solicits distinct but related BOOK concepts. Independent of what a speaker might intend to convey in uttering (17),[15] 'book' is flexible enough to solicit multiple related concepts with a single use.

This flexibility motivates an internalist picture of meanings, whereby the purpose of linguistic meaning is to connect pronunciations with conceptual contents. On this view, meanings are cognitive instructions for retrieving and constructing concepts (Pietroski 2018).[16] Such internalist meanings are not conceptual contents, extensions, or truth conditions—they are protocols to fetch and assemble (new) concepts from concepts stored in the extra-linguistic conceptual system. For ease, we can call such meanings *concept assembly protocols*, or CAPs for short.

An analogy may help. Consider a "color-by-numbers" task, similar to one found in children's coloring books. Success on such a task requires following instructions like "fill numbered regions using crayons from provided boxes" along with details correlating coloring resources with numbered regions—"1 = piros"; "2 = lila"; etc. Given crayons in bins marked 'piros', 'lila', etc., and given identical templates to color, two different children Bia and Cloe might produce different images by executing identical instructions. Depending on the shades they select from the respective boxes, Bia's and Cloe's pictures will bear some clear resemblances, but differ in dramatic ways. If Bia grabs a ruby-colored crayon, and Cloe grabs a brick-colored crayon for shading the 1-areas of their type-identical image outlines, both aptly execute the instruction to shade the 1-numbered regions with a crayon from the piros-box. But they produce clearly distinct products. Despite using the same coloring frame (a flower outline, say) and shared color resources (the Hungarian marked boxes of crayons), Bia and Cloe execute the same instructions to produce different images.

---

[15] For example, a speaker might intend to convey that the location of the physical vehicle is an indication that they disagree with the content's critical reception; or they might intend to convey exactly the opposite, that the vehicle's location should not be taken as a mark of the content's quality. But whatever the speaker might mean/intend, the expression 'the blue book' itself must have a meaning flexible enough for the audience to arrive at the *abstract* referent by way of referring to the *concrete* physical vehicle.

[16] For roughly similar approaches to theorizing about linguistic meaning in general, see for example Carston 2016, 2019; Glanzberg 2011, 2014, 2018; Harris 2020; and Pritchard 2022. While we couch it in the semantics of Pietroski (2018), our basic proposal concerning slurs and their meanings might instead be situated instead within one of these alternatives to CAP. We leave it to future work to establish commonalities and differences among the approaches to slur meanings that might be taken within each of these similar frameworks for semantic theory.

For the CAP internalist, meanings are likewise instructions. A sentence provides a grammatical frame, like an image outline, which is filled with the conceptual contents retrieved in executing instructional meanings. For example, the meaning of the expression 'blue book' is a complex instruction to retrieve two concepts—a BLUE concept and a BOOK concept—and combine them. If we think of various related concepts as stored in a single mental location or address (like crayons stored in labeled boxes), the meaning of a word (like 'book') is an instruction to grab one of the (BOOK) concepts at that location and combine it with the other retrieved (BLUE) concept to accord with the syntactic structure of the sentence in which the expressions are used. And much like Bia and Cloe, who (let's suppose) speak no Hungarian, such instructions can be executed without any regard for the content of the concepts located at the instructed address.

Meanings as CAPs thereby map contents to grammatical positions in sentences—instructions to grab and combine concepts given the sentence's syntactic frame.[17] But just as different children following the same coloring instructions might produce different (but related) images in a color-by-numbers task, two hearers might form different but related thoughts in executing identical instructional meanings. This flexibility enables the CAP internalist to account for such differences, permitting the fetching of different BOOK concepts from a single use of 'book.'

Take for example our two discussed readings of (16).

(16) The blue book is Sona's.

Again, (16) can be used and readily understood to mean that Sona is related in one way or another either to a physical copy of a book or to its abstract content. Executing the instruction to fetch a BOOK concept on the first reading returns a concept for thinking of things as having a location, leaves, typefaces, etc. On the second reading, executing that *same* instruction retrieves a concept for thinking of things that are authored, abstract, have token instances, etc. The word 'book' encodes *one* instruction, but this single instruction can be carried out by retrieving one of *multiple* concepts fetchable at the instructed mental address.

Such flexibility is thereby captured without (over)populating a multitude of meanings for the many flexible expressions in natural languages. After all, one could accommodate the multiple interpretations of sentences like (16) by insisting that we have as many words as we do concepts (or extensions, or truth-conditional contributions). There are two words, both pronounced /bʊk/, each of which expresses a different concept/extension/satisfaction-condition (see, e.g., Katz 1972; Fodor 1998). But this homophonous explanation fails to explain the flexibility exhibited in sentences like (17), which contains only a single instance of 'book' yet regards (at least) two different BOOK concepts. Whichever word 'book' appears in (17), the anaphoric 'it' must abide by the correlative way of thinking about books as either abstract or concrete. Consequently, one of the predicates is invariably unsatisfied. On the CAP proposal there is a single meaning, the instruction to retrieve a BOOK concept and return it for

[17] For reasons that we cannot explore here, extant CAP proposals constrain the kinds of concepts word meanings recruit. In particular, fetchable concepts, generated in the process of word learning, are (mostly) monadic. The constrained homophony of linguistic meaning suggests that the compositional contours of language do not align with the compositionality of thought. Taking fetchable concepts to be re-packaged monadic variants of polyadic concepts accounts for the systematic way in which meanings can be combined, while respecting the otherwise perplexing boundaries on the thoughts a given sentence can(not) express (Pietroski 2010).

combination—a single instruction that gets executed twice in the sentence in (17) given the anaphoric 'it.'[18]

Before returning to slurs, note that instructions (not just CAPs) are neither true nor false. Bia and Cloe might fail to execute their coloring instructions for many reasons: if the instructions are unclear, if they are not given crayons to fetch, if they are inattentive, etc. But there is no sense in which failure or success in executing those instructions is a matter of truth or falsity. Failure to follow an instruction does not render that instruction false. CAPs, as cognitive instructions, are likewise neither true nor false. For this reason, the CAP view denies what Jason Stanley calls the "Fregean spirit" of externalism for linguistic meaning: that "meanings determine the truth-conditions an utterance of [a] sentence [has]" (Stanley 2007, 8).

The CAP theorist might accept that the *result* of executing a complex instructional meaning is a truth-evaluable thought. That is, the CAP proposal is compatible with certain *acts* of speaking being somehow determinative of truth conditions. But the meaning of a sentence, the complex CAP, does not have such a truth condition, even as derived from the constructed thought. Executing the instructional meaning of (16) might yield two (or more) different thoughts in accordance with the BOOK concept fetched. But even assuming that such constructed thoughts have truth-conditions, the indeterminacy involved in executing a CAP will underdetermine which thought is constructed, and thereby fail to determine a truth-condition. The consequence is that *sentences* do not have truth-conditions, even if thoughts (and perhaps speech acts) do. Correlatively, a given word like 'book' will not have an extension, even if the concepts speakers fetch in understanding uses of 'book' have extensions.

To be clear, cases like (13-17) do not constitute a decisive objection to accepting the extensionalist assumptions about meanings. Various inventive strategies have been developed to deal with individual cases for nouns like 'book.'[19] But natural languages are replete with flexibility, which is not confined to a narrow class of nominals. There may well be extensionalist accounts available for explaining the flexibility of some words, but even proponents of the extensionalist assumptions admit that such solutions will differ in kind for verbs (Vicente 2018 2019). Additionally, it's worth pointing out that these cases of flexibility are not the only source of trouble for the extensionalist assumptions about meanings.[20]

Applying the CAP proposal to the meaning of slurs is straightforward. Slurs have meanings that are instructions to retrieve concepts. The use of 'mick' in

---

[18] Contemporary literature takes examples like (17) as illustrating the *co-predication problem*, since *prima facie* incommensurate predicates are jointly applied to a single noun phrase. But the flexibility, or polysemy, of natural language expressions is not quarantined to nouns. Verbs are likewise polysemous. So, while various truth-conditional strategies have been devised to resolve the co-predication problem for nouns, these strategies will not thereby resolve the general problem that polysemy poses for theories that identify meanings with concepts/intensions/extensions/satisfaction-conditions.

[19] Some respond to these worries by indicating that meanings are hybridized concepts that mirror noun polysemy (Vicente 2019), while others take cases of polysemy to show that inferences of natural language speakers are mistaken (Kennedy & Stanley 2009). Still others advance a version of the homophonous move that applies to the structure of lexical entries (King 2018), while others simply accept that some rather bizarre hybrid objects exist (Gotham 2017, 2022). Others claim there is no problem at all (Burge 2003).

[20] Limitations on space make efforts to detail the variety of considerations informing truth-conditional approaches to meaning impractical. But various problems have been raised: concerns about intensionality (Hinzen et al. 2014) and reference failure (Hinzen 2007); computational demands and linguistic composition (Pietroski 2018); the inconsistency of the extensionalist assumptions (Quilty-Dunn 2021); problems of linguistic ontology (Stainton 2006); and implausible ontological commitments (Collins 2009).

(5) All Irishmen are micks.

is an instruction to retrieve a concept at a mental address, where the concept retrieved enables thinking derogatorily about Irish people. And while one might take that derogatory character to be part of the concept fetched (more on this below), the meaning of 'mick' carries no such content, any more than the instruction to fetch a crayon contains a (part of a) crayon.

The CAP view then resolves the NPC problem for pragmatic accounts of slurs by denying the synonymy and co-extensionality of slur/neutral-counterpart pairs. Slurs simply do not have neutral counterparts. That is, the meaning of a slur will be a different CAP (a different instruction) than any purported neural counterpart. The instructional meaning of 'Irishman' is *not* the same as the instructional meaning of 'mick'—they are different instructions to fetch a different class of concepts from different locations in the extra-linguistic conceptual system. Because of this, a slur cannot be coextensive with a proposed non-pejorative counterpart, since meanings *qua* instructions do not determine extensions. Thus, while most pragmatic accounts of slurs seemingly accept that they have neutral counterparts, given an internalist semantics of the sort sketched above, a pragmatic approach need not accept this assumption.

The NPC argument, recall, is that if we deny that a slur's derogatory character is part of its meaning, then its extension should be expressible with a word that lacks the derogatory impact. Such a co-extensional neutral counterpart can be used, says the argument, to make true claims that cannot be made using the slur—after all, no Irishmen are micks. Thus, any pragmatic account seemingly makes a false prediction: that any slur has a truth-conditionally equivalent neutral counterpart.

However, by adopting the CAP proposal that word meanings are instructions, we can offer a pragmatic approach to the derogatory character of slurs that avoids this line of objection altogether. On the internalist CAP view of meanings, slurring expressions do not have meanings with truth-evaluable contents of any sort. There is no non-pejorative truth-conditional content to a slur that a neutral counterpart can likewise express. Derogatory content may be a property of the concept(s) at the mental address that a slur's instructional meaning calls for, but such content is itself neither an instruction nor part of one. As such, a slur has no extension that a purported neutral counterpart can share. This allows the CAP view to maintain that the derogatory character of a slur is not found in its meaning, while denying the claim in (NPC) that slurs have neutral counterparts.

Bia and Cloe might finish their coloring-by-numbers task and create a picture that comports with reality. If those pictures resemble some part of the world, it seems apt enough to think of their depictions as true (or accurate). But the *instructions* they followed to develop that image cannot likewise be said to be true (or accurate). Instructions, as a rule, are neither true nor false. Likewise, the complex instructional meanings of (5) and (6) are neither true nor false.

(5) All Irishmen are micks.

(6) All Irishmen are Irishmen

This is not to deny that speakers distinguish between sentences like (5) and (6), accepting that (6) is somehow trivial and (5) not. They surely do. But on the CAP view that difference does not directly target a feature of natural language meanings. Rather, the difference speakers attend to in shunning (5) but not (6) has to do with the content of the different thoughts that one arrives at by executing the different CAPs encoded by those sentences.

4 The flexibility of the proposed account, and its limitations

The positive CAP proposal addresses a central challenge for pragmatic accounts of slurring expressions, namely the threat posed by the NPC problem. We have left unspecified exactly how to explain the derogatory character of slurring expressions. In what follows we explore three proposals for how a proponent of the CAP account of linguistic meaning might explain the derogatory character of slurs. The three proposals correspond, roughly, to three classes of proposals extant in the literature on slurs: a conceptual option, an expressivist option, and (what we call) a post-instruction-execution option. The goal here is to show that the internalist CAP proposal is consistent with various strategies for explaining the semantic-pragmatic behavior of slurs. We argue that this is one more virtue of the CAP proposal.

4.1 Option 1: Pejorative content

The first proposal identifies the derogatory character of slurs as part of the content of the concept(s) a slur meaning fetches. If the meaning of 'mick' is an instruction to grab a MICK concept, then the derogatory aspect of the use of 'mick' stems from the content of the concept fetched. On this proposal the concept (somehow) contains the offending material, which manifests in the thoughts that speakers and audiences construct in executing the instructional meanings of sentences like 'Conan O'Brien is a mick.' In other words, this proposal accepts the CAP thesis that linguistic meanings are not contents, but instructions, while locating derogatory content in one of the concepts located at the address indicated in the instruction.

This proposal is enticing, since it offers plausible explanations for several phenomena related to the use of slurs. First, it explains why speakers have difficulty canceling the derogatory character of a slur. Because the offending content is constitutive of the concepts used in executing meanings, a speaker cannot merely forestall an unintended inference by her audience in the way one might for (say) a conversational implicature, as in (18).

(18) (a) Matt: Does Sona love Conan?
(b) Tak: Sona likes Conan.

Tak's reply to Matt implicates that Sona does not love Conan. Tak implicates, by using the sentence in (18b), a thought roughly expressed by (19) and indicated more explicitly in (19'):

(19) Sona does not love Conan.

(19') $\exists e \exists x[[\text{AGENT}(e,x) \,\&\, \text{SONA}(x)] \,\&\, \text{NOT-LOVE}(e) \,\&\, \exists y[\text{PATIENT}(e,y) \,\&\, \text{CONAN}(y)]]$

This thought is distinct from one built upon executing the meaning of (18b), namely:

(18b') $\exists e \exists x[[\text{AGENT}(e,x) \,\&\, \text{SONA}(x)] \,\&\, \text{LIKE}(e) \,\&\, \exists y[\text{PATIENT}(e,y) \,\&\, \text{CONAN}(y)]]$

The thought in (18b') contains the concepts fetched in executing the instruction in (18b), viz. LIKE(_), that are absent in (19'). Because these are distinct thoughts constituted by distinct concepts, with (19') derived somehow from (18b') to generate the conversational implicature, as a speaker Tak can indicate that the relevant derivation makes unendorsed assumptions; e.g. "But she also loves him." Tak's authority *qua* speaker has jurisdiction over the aptness of the inference from (18b) to (19), but has limited say over how an audience ought to execute the instructional meaning of (18b) to build (18b').

On a particular utterance, the fetched MICK concept is a constituent of the thought that an audience builds in understanding (20).

(20) Conan O'Brien is a mick.

If the derogatory character of the slur 'mick' is due to (some part of) the content of the fetched MICK concept, no amount of speaker clarification can prise that derogatory content from the

fetched concept. The speaker cannot (easily) strip their utterance of its derogation because the derogatory content is a constituent of the thought the sentence instructs one to build. Where (19') can be canceled because it is distinct from (18b'), there is no additional reasoning that is responsible for the derogatory thought that a particular use of (20) instructs us to build on those occasions where the derogatory MICK concept is fetched.

Second, thinking of derogation as due to conceptual content can also explain how cases of reclamation (or NDNA uses) are possible. While the conditions for a successful reclamation of a slur are difficult to articulate (Cepollaro & López de Sa 2022), presumably those who have reclaimed a given slur have performed some cognitive operation that others, at least prior to reclamation, have not. When reclamation is effected by the group a slur targets, group members have come to identify themselves with uses of the slur, coming to think of themselves as falling under the related concept. But of course, they do so without the negative aspect historically attached to the slur. For example, in the case of the slur 'queer' this reclamation came about as a means of unifying the experiences of those marginalized because of their sexual orientation (see, e.g., Bérubé & Escoffier 1991, 14).[21] Individual members of these communities came to *think of themselves* as queers in a way that is empowering and unifying, not degrading. One way of explaining this reclamation is that, at the cognitive level, individuals in queer communities each developed a new (shared) concept QUEER, distinct from the one previously used to think about members of these marginalized groups in derogatory ways. Like the derogatory concept, this empowering concept is retrievable with uses of 'queer.' If the derogatory character of a use of 'queer' stems from the content of a (negative) concept QUEER$^-$, the process of reclamation plausibly involves developing a concept QUEER$^+$ that strips the pejorative content from the concept, replacing it with positively valenced content.

This change might manifest cognitively in multiple ways. Here, we suggest two. The first holds that 'queer' is polysemous (Ritchie 2017; Jeshion 2020). On this proposal 'queer' (much like 'book') is capable of expressing multiple concepts—a derogatory one and an empowering one. Reclamation is the process of developing an additional QUEER$^+$ concept, fetchable alongside the historically derogatory concept QUEER$^-$ with the single expression 'queer.'

This would explain why some special contexts permit the "canceling" of the derogatory character of a slur's use. Much the way certain situations influence the fetching of a BOOK-vehicle concept, and not a BOOK-content concept ('The book weighed exactly 1kg' vs. 'The book was plagiarized'), particular situations influence the fetching of the empowering QUEER$^+$ concept. For the case of reclamation, one obvious factor that influences which concept is fetched would be the membership of the speaker in the community the slur traditionally targets (see Kennedy 2002, Ch. 3), which might likewise explain the differences between in-group and out-group uses.

A second strategy treats the process of reclamation as adding to the lexicon, developing an additional word that shares a (similar) pronunciation with a slur. In the case of 'queer' the claim would be that we have developed a second word 'queer$_2$' that fetches the empowering concept QUEER$^+$, and not the historically derogatory concept fetched by the original slur 'queer$_1$' (see also Jeshion 2013 on reclamation via homonymy).

[21] As the authors explain, reclaimed uses of 'queer' are "meant to be confrontational—opposed to gay assimilationists and straight oppressors while inclusive of people who have been marginalized by anyone in power… to combine contradictory impulses: to bring together people who have been made to feel perverse, queer, odd, outcast, different, and deviant, and to affirm sameness by defining a common identity on the fringes" (Bérubé & Escoffier 1991, 14).

These two proposals appeal to different cognitive structures in accounting for the reclamation of slurs within an internalist framework. Each structure may be apt for different cases of apparent reclamation. The first posits one lexical entry that connects to a single address containing multiple concepts. The second posits two lexical entries, and two distinct addresses that each contain a distinct concept. One advantage of the internalist proposal is that it can accommodate both accounts, while maintaining a univocal approach to linguistic meaning. While we do not wish to argue here that a pluralist account of reclamation is required, one virtue of the CAP proposal is its compatibility with such a pluralism. For a given slur, parsimony demands a singular account. But across the class of slurs, multiple accounts might be preferable. Because slurs constitute a class of expressions that exhibit complex and conflicting properties, a complete accounting for that variability may require positing different explanations in various cases. To see this, consider the difference between two courses of reclamation, that for 'queer', on the one hand, and the reclamation of a well-known derogatory term for Black Americans in English. For the latter, reclamation arguably resulted in coining of a second word, with a different pronunciation and orthography compared to the original slur.[22] Insofar as phono-syntactic properties are essential to identifying lexical entries, the homophony account (at least partly) explains differing reactions to the two words, the slur and the "reclaimed" coinage derived from it (Kennedy 2002, 5). In contrast, the word 'queer' seems to have retained its phonological and orthographic properties in the process of reclamation. If one takes this apparent contrast to suggest that a single word connects distinct but related concepts in the case of 'queer,' whereas the n-word exhibits something like a new coinage, one might distinguish two phenomena that are labelled with "reclamation." One of these is accurately described as such, as in the case of 'queer,' where an existing word is put to a non-derogatory use. Other cases appear to involve not the *reclamation* of an existing word by the historically derogated group, but *creation* of a new word.

These considerations alone are not conclusive about the most apt account for these particular slurs, if for no other reason than proving that a particular term is polysemous can be a slippery business (Sennet 2016). It is quite possible that the best theory of reclaimed uses will posit a unified mechanism. But whatever the details about these two slurs and their reclamation, the meanings of these expressions can be given a univocal treatment under the CAP proposal—they are instructions to fetch concepts. On the other hand, the CAP proposal can also be paired with a univocal account of reclamation and its cognitive routes. The theoretical upshot of taking meanings to be instructions is that it offers multiple strategies in accounting for the relationship between words and their corresponding concepts, delivering a flexibility that reflects the complexity of slurs as a class.

4.2 Option 2: Pejorative association

If the derogatory character of a slur is not attributed to conceptual content, one might think it stems from an association between the concept(s) fetchable at the instructed address and another negative mental state or attitude. On this expressivist proposal the derogatory character of slurs is not descriptive (or truth-conditional) content, neither at the level of linguistic meaning nor at the level of the thought constructed in accordance with linguistic meaning. Instead, derogation is somehow a reflection of a speaker's negative attitude toward the group the slur

[22] Tupac Shakur marks the importance of this alternate spelling in a lyric, suggesting that the reclaimed word has significance as an acronym: "I'm Never Ignorant, Getting Goals Accomplished" (Shakur et al. 1991).

targets. This explains why slurs tend to "scope out" in various grammatical frames. For example, in belief reports, a use of a slur tarnishes the speaker of the report, not merely the subject:

(21) Matt thinks that that mick Conan should give Sona a raise.

The use of 'mick' here does not guard the speaker in (21) from disapprobation, even if they are accurately reporting the content of Matt's derogatory mental state by using the very words that Matt used in the original reported utterance. Similarly, the derogatory character of a slur survives negation:

(22) Matt: Liza is dating that lanky mick.
Sona: No, she's not [dating a mick], he's an American.

Much like (21), Sona in (22) is met with the same disapprobation as Matt, even though the slur is under the scope of negation, which tends to *deny* the content of the negated expression. The expressivist account can capture these phenomena because the derogatory character of the slur is not part of the negated content of the fetched MICK concept, but stems from the association of this concept with a negative attitude or emotion.

This additional content that Potts (2007) calls the "ineffable" aspect of slurs is, on this expressivist option, like the content of expletives such as 'damn.' For example, in (23) the use of 'damn' does not seem to predicate any truth-evaluable content of the indicated insects.

(23) Those damn ants are back.

(24) Those poisonous ants are back.

The ants in (23) have no property that would satisfy the damning predicate. In contrast, the ants in (24) are said to have a property, presumably of causing a detrimental biological response in bitten humans. The use of 'damn,' it seems, indicates that the speaker has a particular attitude toward the subject, rather than predicating any sort of property of the object that would satisfy some descriptive content.

For expressivists, the derogatory character of the use of slurs is identified with a similar "ineffable" content. For expressivists who endorse our CAP proposal, this might be the result of associations that the speaker (and audience) have with the particular concept retrieved in executing the instructional meaning of the slur. So, the difference between the concepts MICK and IRISHMAN is not a difference in (truth-evaluable) satisfaction conditions, but rather with associations speakers/hearers manifest in their cognitive architecture that tie attitudes/emotions to the deployment of the fetched concept.

One potential advantage of this view is that it offers a straightforward explanation for the derogatory use of *prima facie* non-derogatory expressions.

(25) Be quiet child[/woman/employee].

The general nouns in (25) can be used in ways that derogate the addressee *qua* member of the indicated group. But, of course, being (say) a child does not typically indicate that someone should be targeted for derision, in the way being called a 'mick' does. The expressivist proposal explains these data, since on such uses, speakers betray negative attitudes they hold towards the indicated group, even when the expression has no history of derogation. Setting aside the merits of an expressivist account, our goal here is merely to sketch how the expressivist approach can be couched within the CAP framework.

4.3 Option 3: Post-instruction-execution

The final option is a broad class of possible proposals about the derogatory character of slurs appealing to features of language use, and particularly the use of a slur on a given occasion. Expressed in terms of the CAP view, this cluster of proposals takes the derogatory character of

the use of a slur as stemming from reasoning about, or responding to, the thought one builds in understanding a slurring sentence on the particular occasion of use. Because this class of views takes as an input the completed thought that meanings are instructions to build, we call this class *post-instruction-execution* views. For example, given an instruction to build some thought that utilizes a MICK concept and not an IRISHMAN concept, one might reason that the speaker's instruction to deploy a MICK concept as opposed to some other concept indicates that the speaker takes individuals that fall under the MICK concept to be deplorable—otherwise, the speaker would have offered a different instruction.

This simplified Gricean picture is but one post-instruction-execution option available to a CAP theorist. The CAP view is consistent with any number of such views regarding the derogatory nature of slur uses. Accounts that explain these uses as presuppositional (e.g., Schlenker 2007), a matter of conversational implicature (e.g., Camp 2018), conventional implicature (e.g., Potts 2007, 2012; Williamson 2009; McCready 2010), or as a broader convention or practice (Anderson and Lepore 2013; Hess 2021) are all easily compatible with a view of natural language meanings as instructions to build concepts.

We've reviewed three general strategies for explaining the derogatory character of slurs, couched in terms of the CAP view. While these different views may have various advantages and drawbacks, our purpose here is not to adjudicate among them, or even to unduly trumpet them over competitors. Rather, we simply indicate that any one of them is consistent with a CAP view of natural language meanings. Moreover, the compatibility of the CAP proposal with the various accounts rehearsed above may be necessary in accounting for the diverse behavior of slurs. We have not argued that a pluralist approach to slur derogation is necessary, whereby different accounts are needed to explain the diverse semantic/pragmatic properties of slurs. However, if a complete account of the derogatory nature of slurs requires different approaches for different subclasses of slurs, the flexibility of the CAP proposal is all the more attractive. Since prejudice and oppression are complex social phenomena, the derogatory character of slurs that is closely tied to them may need to be explained by several theories. We may need to accept that slurs can be reclaimed (say) because they are polysemous, yet are offensive across various grammatical frames because their use (say) often violates social norms. If this is right, then a purely semantic account of the derogation of slurs, one that hopes to unify these diverse phenomena under a univocal explanation, is fundamentally flawed (DiFranco 2015). In contrast, if meanings are instructions to fetch and combine concepts, this permits a wide variety of pragmatic mechanisms to explain the multi-faceted properties of slurs and their use.

5 Distinguishing CAP from some alternative views

In this section we address three alternative approaches that bear important similarities to our view. We have two main aims in doing so: first, to illustrate the novelty of the CAP proposal in the literature on slurs; second, to point to considerations that are of central importance in adjudicating among the options. As will be seen, the choice among these approaches to derogation depends on how we settle foundational questions about cognition—for example, about the architecture of our conceptual system and the degree of complexity in the information encoded by words.

In several papers addressing different subclasses of slurring expressions, Croom (2011, 2014a, 2015a, 2015b, 2018) defends a hybrid approach on which slurs have a two-component semantics. A given slur, such as 'mick,' has a meaning with both a descriptive and an expressive component, where the descriptive component includes a list of weighted prototypical features,

some positive and others negative. The extension of the predicate 'mick' consists of individuals who are related to one another not via a shared essence, but via a relation of family resemblance. So, there is no one feature or subset of features that all members of the category exhibit, with different members instead sharing different subsets. Furthermore, the properties in the list are ranked in a way that is determined contextually—specifically, by the speaker's selection, from among the listed properties, those which are most appropriate for the conversation at hand. Derogatory uses result from selection of negatively valenced properties, whereas insulated uses either involve selection of positively valenced features or reevaluation of negatively valenced features. On Croom's account, since the features encoded in the lexicon for 'mick' differ from those encoded in the lexicon for 'Irishman,' the expressions are not truth-conditionally equivalent, differing in the extension determining properties that are contextually ranked on a particular use. Likewise, for any given slur, the prototypical features encoded in its descriptive component will differ from those encoded by any putative non-slurring counterpart. Hence, Croom's view avoids the NPC problem.

There are two key differences between this view and ours. First, Croom's account of a slur's derogatory character is semantic, locating derogation in a slur's lexical entry in the form of negatively valenced prototypical feature representations. By contrast, on the CAP proposal, the semantic features of a lexical entry are exhausted by the instruction to fetch a concept at a mental address, which includes nothing derogatory. The contrast is subtle but important. On Croom's view, pragmatic mechanisms play a crucial role, determining which prototypes/properties in a word's lexical entry are given primacy on a particular use, rendering that use either derogatory or not. This places substantive constraints on the kind of pragmatic mechanism at work in understanding a slur utterance, which must be both sophisticated enough to rank some prototypes /properties over others, yet simple enough so as not to require a completed truth-evaluable thought as an input. After all, on Croom's view the truth-conditional contribution of a slur is determined contextually by the process that ranks such features. In contrast, while the CAP proposal allows that some of the concepts housed at the address indicated by a slur's meaning may themselves be derogatory, the account posits nothing derogatory in the lexicon—in the stored meaning of the slur—itself. Accordingly, the CAP view is more firmly committed to a pragmatic explanation of derogation, while placing fewer constraints on the intricacy of the relevant pragmatic mechanism. Second, Croom's account posits a richly or "thickly" structured lexicon, whereas the CAP proposal is compatible (and traditionally paired) with a "thin" semantics, on which words having as their meanings, not lists of descriptions, but highly abstract and under-specific representations (viz., an instruction to collect a concept at an address). As far as we know, ours is the first defense of a pragmatic approach to derogatory character that responds to the NPC challenge by leveraging the virtues of a thin semantics. Like Croom's, the two other views with which we compare the CAP proposal in the rest of this section similarly involve appeal to rich lexical meanings.

Zeman (2022) applies to analysis of slurring expressions the semantic framework of Del Pinal (2018), a development of Pustejovsky's (1995) qualia structure semantics. On this approach, a word encodes several lists of features grouped together under different "meaning dimensions," with different word classes (e.g., nouns versus verbs) encoding features along different dimensions. So, for example, a noun like 'book' might encode *perceptual* features such as size, shape, color; *constitutive* features such as having a cover or having pages; *agentive* features such as having been written or printed; and *telic* features like being meant to entertain. This approach explains polysemy in general by appeal to foregrounding and backgrounding of

different meaning dimensions in different contexts. Zeman's proposal is to give such a semantics for slurs, using as a central datapoint the Romanian slur 'țigan' (roughly translated as 'gypsy).' Zeman suggests that this slur might encode features in five meaning dimensions, having in common with 'book' the features *perceptual*, *constitutive*, and *agentive*, but without *telic* and with the addition of *evaluative* and *origin*. The *evaluative* dimension encodes a negative evaluation that is based on a range of stereotypes, while *origin* specifies things like geographical provenance, history, and social status. Different uses of 'țigan' highlight different meaning dimensions via a combination of foregrounding and backgrounding. Derogatory uses, as in G-referencing, -extending, and -contracting uses, all foreground the *evaluative* dimension. Appropriated uses of a slur likewise involve foregrounding of the *evaluative* dimension, but with the valence switched from negative to positive. In contrast, non-derogatory, identificatory uses of 'țigan' (in which Roma use the term in self-reference and in reference to other Roma) background the *evaluative* dimension while foregrounding *origin*. This is only a brief sketch of a qualia structured approach like Zeman's but will suffice for the purposes of individuating our own proposal in the space of existing views. Zeman's (2022) theory of slur meanings avoids the NPC problem by taking slurs to be polysemous. Not only will the features encoded by a non-slurring expression/counterpart differ from those encoded by a slur, ensuring truth-conditional non-equivalence, but the resolution of polysemy on a given use will select different features of the polysemous slur, ensuring that no proposed NPC will be coextensive.

Finally, on Neufeld's (2019) rich-lexicon approach to slurs, a term like 'mick' encodes three things: a representation of a group essence, a set of negative stereotypes associated with the group, and a representation of a causal link between the essence and the stereotypes. The first of these three parts is familiar from externalist accounts of natural kind terms like 'water' (e.g., Putnam 1975). To explain derogatory uses of slurs, Neufeld argues that the lexicon must encode negatively valenced features that are causally linked to a group essence. Importantly for slurs, targeted groups lack the encoded essence, making slurring expressions *failed* kind terms (compare 'phlogiston'). Neufeld's theory, like those of Croom and Zeman, avoids the NPC problem. Neufeld maintains that a non-slurring race term like 'Irishmen' likewise encodes a representation of group essence together with causal links to stereotypes both negative and positive. But for non-slurs these stereotypes are "much more innocuous" and the represented causal link between essence and stereotypes are weaker than in the case of a slur (2019, 5). Hence, 'mick' and 'Irishman' are not synonymous, and so the NPC problem is avoided on this third theory as well.

These three views illustrate one strategy for dealing with the NPC problem. The solution in each case involves appeal to a rich lexicon. If a slur's meaning is a richly structured object consisting of various representations—whether Croom's list of prototypes, Zeman's set of feature lists, or Neufeld's essence-imputing stereotypes—the NPC problem is avoided. The CAP proposal, by contrast, avoids the NPC problem without postulating any such complexity in the lexicon or the representations that word meanings express. The meaning of 'mick' on the CAP approach is comparatively "thin" as an instruction to fetch a concept at an address. Hence the CAP proposal offers several pathways for explaining a slur's derogatory character that also answers the challenge of the NPC problem. So, while the views discussed in this section resemble the CAP proposal in taking word meanings to variably select distinct but related contents in different conversational situations, the CAP proposal is distinct from them. The richness of lexical structures is a live debate, and space constraints forestall any attempt to detail the considerations informing that debate, much less convincingly defend the kind of thin

semantics offered by the CAP proposal.[23] Suffice it to say, a view on which the lexicon encodes very little information remains among the live options in debates about meanings and concepts. But because such a view is often offered to bolster an externalist semantics (e.g. Cappelen & Lepore 2005; Borg 2004, 2012) applying an internalist, "thin" semantic framework to slur meanings in particular distinguishes the CAP proposal from the initially similar seeming views discussed above.

Of the three views we have described in this section, none falls squarely on the pragmatic side of the semantic/pragmatic divide. Neufeld's essentialist account of a slur's derogatory character is clearly semantic, with 'mick' encoding a representation of negatively valenced stereotypes, a representation of a group essence, and a causal link between them. The views of Zeman and Croom are harder to categorize as semantic or pragmatic. In Zeman's lexical entry for a slur, the *evaluative* dimension for 'mick' would include representations of a derogatory attitude toward the target group, and so the view would count as semantic in the way that we have been using the terms. On the other hand, in some contexts of utterance the *evaluative* dimension is backgrounded at the same time as other dimensions are foregrounded, such that pragmatic mechanisms result in a non-derogatory use. Croom's explanation of a slur's derogatory character is likewise partly semantic and partly pragmatic. However, in both cases, derogatory information is encoded in the slur's lexical entry, and so counts as semantic in the way we have been using the terms. By contrast, our CAP proposal is unique in that it answers the challenge of the NPC problem with a decidedly pragmatic account of derogation, and without requiring adoption of a rich/thick lexicon.

So, in effect, we have argued for two conditional statements in this paper. Although we are optimistic about the prospects of future work effectively arguing for a chain of modus ponens inferences using these conditionals, our aim here has been to establish them on their own. To recap, the first conditional says that, if a pragmatic approach to a slur's derogatory character is preferred over a semantic one, then the challenge raised by the NPC problem needs to be addressed, without belying the core pragmatic commitment to explain derogation by appeal to features of use rather than encoded linguistic meaning. Our CAP proposal offers, very clearly within a pragmatic framework, a novel approach to explaining a slur's derogatory character that resolves the NPC problem. According to our second conditional claim, insofar as there are compelling objections against a rich-lexicon approach to word meaning, our CAP proposal offers an attractive approach to explaining slur's derogatory character that avoids those objections while addressing the NPC problem.

6 Objections and replies

The CAP proposal's compatibility with a wide range of explanations of the derogatory character of slurs (as discussed in Section 4) might be seen as a shortcoming. One might object that the CAP proposal itself offers no clear guidance in accounting for the linguistic behavior of

[23] See Glanzberg 2014 for a helpful discussion. See Hogeweg & Vicente 2020 for a defense of "rich" lexicon views. See also Vicente 2018 for defense of a hybrid view on which some words encode rich meanings, while others encode thin meanings. Zeman adopts a rich approach to the lexicon without offering an argument in favor of this framework over "thinner" alternatives (2022, 954). Neufeld similarly offers no argument for a rich approach to the lexicon in general, arguing only that her "background semantic framework…is a live option in current debates about the nature of meaning and conceptual structure" after responding briefly to one common objection (2019, 7). We take the same strategy here, adopting a thin approach to the lexicon without offering extended argumentation for that approach, since our goal is to illustrate what is unique and attractive about the CAP approach to a slur's derogatory character.

slurs. However, this objection rests on controversial expectations about the kind of explanatory work a narrowly linguistic theory should do. For one, the view offers a clear answer as to where we should *not* look for a solution to the problems that slurs pose—according to the CAP proposal, linguistic meanings do not encode the kind of derogatory content that semantic accounts of slurs require. Second, the objection assumes that the various features of slurs can be explained by a single pragmatic mechanism. We've suggested above that this assumption may be incorrect given the relationship between slurs and the complex systems of oppression they reify.

A seemingly more pressing worry is that the CAP proposal does not actually resolve the NPC problem. After all, any theory of slurs will, as indicated in §2, need to account for speaker judgments that sentences like (6) are seen as trivially true, while sentences like (5) are not.

(5) All Irishmen are micks.

(6) All Irishmen are Irishmen.

The CAP proposal purports to explain such basic facts by indicating that these judgments do not reflect the properties of sentences. While sentences are not truth-evaluable, because their meanings are instructions, speaker judgments about truth and falsity derive from the truth-evaluability of *thoughts*. But now, the objection insists, the NPC problem arises at the level of thought. Whatever the formal details, the two thoughts that (5) and (6) instruct us to build differ in content because of a difference in constituency:

(5*) … IRISHMEN( ) … MICK( )

(6*) … IRISHMEN( ) … IRISHMEN( )

The thought in (6*) is trivially true because the same concept is fetched on both executions of the instructional meaning of 'Irishmen', while (5*) contains distinct concepts, with distinct contents. If speaker judgments about the falsity of (5) are explained by appealing to truth-evaluable thoughts, then the falsity of (5*) seemingly must be due to the difference in the satisfaction conditions of IRISHMEN and MICK—viz. the concepts have different extensions. But now, claims the objection, the difference between slurs like 'mick' and their purported NPCs like 'Irishmen' are articulated in terms of the truth-conditional contributions of fetched concepts. This would preclude most pragmatic views about slurs. The only pragmatic features that can be relevant in accounting for the difference between slurs and NPCs are those that influence the retrieval of concepts. But those factors settle which concept is retrieved in the course of linguistic cognition, before any pragmatic reasoning can take hold. Thus, the derogatory character of a slur cannot be attributed to pragmatic features of a slur's use, beyond those aspects of use that direct concept retrieval.

There are three replies to this objection available to the CAP view. First, the CAP theorist can reject the initial assumption about how to explain the truth-value judgments of speakers. That is, the CAP theorist might deny that the concepts fetched during linguistic cognition have extensions. In presenting the CAP proposal we alluded to a constraint on the kinds of concepts that meanings are instructions to fetch: such concepts must be monadic (see note 17). The considerations about the compositionality of language that motivate the view that fetchable concepts are repackaged, super-dyadic concepts likewise mark the polyadic expressions in the language of thought as distinct in kind, and in truth-evaluability, from the massively monadic concepts that natural language meanings are instructions to build (Pietroski 2018, 230, 259-263, 305). If the concepts fetched in executing the instructions for 'mick' and 'Irishmen' lack extensions, then the fetched concepts cannot have different extensions, and so the objection is avoided.

However, a CAP theorist need not adopt such a restrictive view of fetchable concepts. Even if a CAP theorist assumes that fetched concepts have extensions, a second reply is available to the objection that the NPC problem arises at the level of thought. The CAP theorist might admit that a slur-concept and an NPC-concept are coextensive, yet indicate that users of the relevant concepts simply fail to recognize this. One might well have facility with both a RENATE concept and a CORDATE concept but fail to realize that these concepts apply to the same individuals. This kind of move is objectionable if one collapses the distinction between linguistic and conceptual competence. If what one understands when one grasps the meaning of an expression is what the world would be like if that expression were true/satisfied, one cannot then claim that speakers *have* such a competence and then offer an error theory indicating that they *lack* such competence in a narrow domain of speaker judgments without thereby denying that speakers understand expressions in that class. But the correlative claim about concepts is far less pressing. One can be confused about the things a concept applies to, while still demonstrating facility with the linguistic expression for which such a concept is fetchable. An English speaker can have a competence with 'jade' that is indistinguishable from that of other English speakers who are privy to discussions of jadeite and nephrite. Likewise, a French speaker can demonstrate the same facility with 'Londres' as his fellow Parisians, yet still be confused about the satisfier of a LONDON-concept. Of course, if one collapses the distinction between the satisfaction conditions of concepts and the meanings of natural language expressions, these pedestrian observations lead to implausible claims about speaker competence. But precisely because the CAP theorist denies that meanings determine truth-conditions, they can maintain that speakers understand the meanings of expressions even if they might be confused about which worldly things their concepts apply to.

A third reply to the objection presses on Specificity, which on the objection insists that a slur's group-targeting feature is explained by the fact that the content of slur-concepts would have members of those groups in their extension in the absence of derogatory content. However, all the CAP view requires to resolve the NPC problem at the level of thought is that the slur-concept fetched does not share an extension with the purported NPC-concept. That the slurring *expression* has the feature of Specificity need not follow from the conceptual content an expression fetches. Part of our (non-bigoted) understanding of slurs is that *others* treat them as applying to groups. While that understanding of slurs seems related to their ability to derogate, this feature need not be a function of extension-determining conceptual content. To make this a bit more explicit, consider what the content of a MICK concept might be like. Such content could be purely descriptive, indicating the contours of a stereotype, of the sort one might see in a political cartoon—a description no human satisfies. With such purely descriptive content, the derogation of any use of 'mick' can be the consequence of a speaker taking this MICK-stereotype to apply, not only to an individual, but as applying to anyone that also satisfies IRISHMEN. In terms of conceptual content, MICK and IRISHMEN differ, and yet neither has any derogatory descriptive content. Rather, the derogation stems from the use of the slur, which offensively implies that the stereotyped MICK-concept applies everywhere IRISHMEN applies. So, while a MICK concept might well contain the descriptive content of an IRISHMEN concept, that difference in content need not be derogatory. As such, IRISHMEN need not simply be MICK, stripped of its derogatory descriptive content. Rather, IRISHMEN is what is left when the descriptive content of MICK that, *when implicated as applicable to all Irishmen* constitutes a derogatory *utterance,* is stripped away. Importantly, on the CAP proposal, such an avenue is available without denying that we use words to say the same thing as one another. If other authors (e.g., Cappelen & Lepore

2005, 126) are correct that two speakers can say the same thing in uttering a sentence, this Fregean commitment is compatible with the CAP proposal. When the bigot uses 'mick' to deride Conan, he indicates the exact same instruction that a history professor discussing racist institutions does in using 'mick.' Both speakers say the *same thing*, in verbalizing the same instruction: <<fetch a MICK concept>>.

A third and final objection appeals to parsimony. Here the worry is that, despite the dilemma the NPC problem poses, surely this problem cannot justify upending the orthodoxy of externalism. On this criticism, both the fecundity of truth-conditional semantics, and the comparative paucity of internalist explanations for basic phenomena together demand that we not reject externalism without a good deal of trepidation. The NPC problem, on this objection, does not meet that benchmark. Stanley (2007) makes a similar point more broadly, indicating that internalist proposals fail in principle where externalism promises success. If we pessimistically reject externalism "it is mysterious how language-users could so smoothly move from linguistic comprehension to action… [I]f the pessimistic [i.e., internalist] view were correct, the connection between speech and action would be inexplicable" (Stanley 2007, 9). He continues:

> Furthermore, if the pessimistic [i.e., internalist] view were correct, it would not be clear how to construct a theory of meaning. Native speakers have intuitions about the truth and falsity of what is said by an utterance, relative to various possible situations. But native speakers do not have intuitions about meaning… In short, [internalist] theorists of meaning who do not think there are systematic relations between the intuitions native speakers have about the truth and falsity of utterances and the meanings of words and sentences have stripped themselves of any plausible evidential basis for their hypotheses. (Stanley 2007, 9)

As such, while denying that meanings determine truth-conditions plausibly resolves the NPC dilemma, such a denial comes at a substantial cost, namely the benefits and progress wrought by the externalist research program, which presumably can both explain the connection between linguistic competence and behavior, and can do so by taking speaker judgments about the truth-value of utterances as evidence for theories that offer such explanations.

This objection assumes the externalist program has a laudable history of success in explaining natural language speaker data. One might, however, suspect both that the past achievements have been exaggerated (Pietroski 2015), and that externalia in fact do not play a central role in the proposed explanations (Collins 2021). But setting aside such skepticism, the utility of appealing to the truth-evaluability of utterances can plainly be preserved on the CAP proposal, *pace* Stanley. For all we have said about meanings as lacking truth-conditions, the CAP view is perfectly consistent with the view that the thoughts meanings are instructions to build are themselves truth-evaluable. As such, speakers' truth-value judgments about utterances, as judgments about the thoughts that meanings are instructions to build, offer insight into how meanings are related to the information the world provides (for example, regarding quantification (Knowlton et al. 2022, Lidz et al. 2011) and event structure (Wellwood et al. 2018)). Likewise, utterances on such a CAP view offer instructions for building truth-evaluable thoughts which convey information to their audience—thoughts that make contact with the world, offering the very kind of explanation that Stanley claims only externalism can provide. An assertive utterance containing the word 'chocolate,' used to tell a child about where in the kitchen it might be found, offers an instruction to build a thought containing a CHOCOLATE

concept used for thinking about things as such. There is no more mystery as to how such a thought can convey information about a candy bar's location than there is for the externalist explanation in which a sentence/utterance conveys that information to the child. Hence, the claimed unique virtues of externalism in explaining why the child would bound off to the kitchen as a result of hearing the relevant utterance can be retained by the CAP proposal.

7 Conclusion

The derogatory character of a slur appears to be multi-faceted, shifting in ways that require a flexibility that many truth-conditional accounts struggle to explain. The main argument in favor of truth-conditional semantic accounts claims that these difficulties are outpaced by the problem posed by non-pejorative counterparts (NPC), which pragmatic accounts presumably cannot resolve. We've shown that this parsimony is illusory, since the NPC problem does not follow from pragmatic theoretical commitments regarding slurs. Rather, an independent commitment—that the meanings of expressions determine their extensions—is the source of the NPC problem. Not only can a pragmatic account of slurs deny this more general semantic claim in favor of an *internalist* semantics, but there are independent reasons to do so. And the internalist framework on offer, that treats meanings as CAPS, not only forestalls the NPC problem, but also motivates a multifaceted pragmatic account of slur behavior that can help explain the complex ways in which speakers can slur.

Acknowledgements

For helpful feedback on earlier versions of this manuscript, the authors would like to thank Aiden Woodcock, Alexander Williams, attendees at a 2022 symposium at Cambridge University, and attendees at the 22nd Szklarska Poręba Workshop on the Roots of Pragmasemantics.

References

Anderson, Luvell 2018. "Calling, Addressing, and Appropriation." In: *Bad Words*: *Philosophical Perspectives on Slurs*. Edited by David Sosa, New York: Oxford University Press, 6-26.

Anderson, Luvell and Lepore, Ernie 2013. "Slurring Words." *Noûs* 47(1), 25-48.

Aristotle 1984. "Categories." In: *The Complete Works of Aristotle*, Vol. 1. Edited by Jonathan Barnes, Princeton: Princeton University Press, 3-24.

Asher, Nicholas 2011. *Lexical Meaning in Context: A Web of Words*. Cambridge University Press.

Ashwell, Lauren 2016. "Gendered Slurs." *Social Theory and Practice* 42(2), 228-239.

Austin, John L. 1961. *Philosophical Papers*. Oxford: Oxford University Press.

Bérubé, Allan and Escoffier, Jeffrey 1991. "Queer/Nation." *Out/Look: National Lesbian and Gay Quarterly* 11, 12-23.

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

Bolinger, Renée Jorgensen 2017. "The Pragmatics of Slurs." *Noûs* 51(3), 439-462.

Bolinger, Renée Jorgensen 2020. "Contested Slurs: Delimiting the Linguistic Community." *Grazer Philosophische Studien* 97(1), 11-30.

Borg, Emma 2004. *Minimal Semantics*. Oxford: Oxford University Press.

Borg, Emma 2012. *Pursuing Meaning*. Oxford: Oxford University Press.

Blakemore, Diane 2015. "Slurs and Expletives: A Case Against a General Account of Expressive Meaning." *Language Sciences* 52, 22-35.

Brontsema, Robin 2004. “A Queer Revolution: Reconceptualizing the Debate over Linguistic Reclamation.” *Colorado Research in Linguistics* 17, 1-17.
Burge, Tyler 2003. “Psychology and the Environment: Reply to Chomsky.” In: *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Edited by Martin Hahn, and Bjørn Ramberg, Cambridge, MA: MIT Press, 451-470.
Camp, Elisabeth 2013. “Slurring Perspectives.” *Analytic Philosophy* 54(3), 330-349.
Camp, Elisabeth 2018. “A Dual Act Analysis of Slurs.” In: *Bad Words: Philosophical Perspectives on Slurs*. Edited by David Sosa, New York: Oxford University Press, 29-59.
Cappelen, Herman and Lepore, Ernie 2005. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*, Malden, MA: Blackwell Publishing.
Carston, Robyn 2016. “The Heterogeneity of Procedural Meaning.” *Lingua* (175–176), 154-166.
Carston, Robyn 2019. “Ad Hoc Concepts, Polysemy and the Lexicon.” In: *Relevance: Pragmatics and Interpretation*. Edited by Kate Scott, Billy Clark and Robyn Carston, Cambridge: Cambridge University Press, 150–162.
Cepollaro, Bianca 2015. “In Defense of a Presuppositional Account of Slurs.” *Language Sciences* 52, 36-45.
Cepollaro, Bianca 2017. “The Semantics and Pragmatics of Slurs and Thick Terms.” (Doctoral Thesis). Philosophy, Université Paris Sciences et Lettres; Scuola Normale Superiore.
Cepollaro, Bianca and López de Sa, Dan 2022. “Who Reclaims Slurs?” *Pacific Philosophical Quarterly* 103(3), 606-619.
Cepollaro, Bianca, Sulpizio, Simone and Bianchi, Claudia 2019. “How Bad Is it to Report a Slur? An Empirical Investigation.” *Journal of Pragmatics* 146, 22-32.
Cepollaro, Bianca and Thommen, Tristan 2019. “What’s Wrong with Truth-Conditional Accounts of Slurs.” *Linguistics and Philosophy* 42(4), 333-347.
Chomsky, Noam 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
Chomsky, Noam 1977. *Essays on Form and Interpretation*, New York, NY: Elsevier.
Collins, John 2009. “Methodology, not Metaphysics: Against Semantic Internalism.” *Proceedings of the Aristotelian Society*, *Supplementary Volume* 83, 53-69.
Collins, John 2021. “Internalist Perspectives on Language.” In: *Cambridge Handbook of the Philosophy of Language*. Edited by Piot Stalmaszczyk, Cambridge: Cambridge University Press, 157-173.
Croom, Adam 2011. “Slurs.” *Language Sciences* 33(3), 343–358.
Croom, Adam 2013. “How to Do Things with Slurs: Studies in the Way of Derogatory Words.” *Language & Communication* 33, 177–204.
Croom, Adam 2014a. “Remarks on “The Semantics of Racial Slurs.” *Linguistic and Philosophical Investigations* 13, 11-32.
Croom, Adam 2014b. “Spanish Slurs and Stereotypes for Mexican-Americans in the USA: A Context-Sensitive Account of Derogation and Appropriation.” *Pragmática Sociocultural / Sociocultural Pragmatics* 8(2), 145–179. https://doi.org/10.1515/soprag-2014-0007.
Croom, Adam 2015a. “The Semantics of Slurs: A Refutation of Coreferentialism.” *Ampersand* 2, 30–38.
Croom, Adam 2015b. “Slurs and Stereotypes for Italian Americans: A Context-Sensitive Account of Derogation and Appropriation.” *Journal of Pragmatics* 81, 36-51.
Croom, Adam 2018. “Asian Slurs and Stereotypes in the USA. A Context-Sensitive Account of Derogation and Appropriation.” *Pragmatics and Society* 9, 495-517.

Davis, Christopher and McCready, Elin 2020. “The Instability of Slurs.” *Grazer Philosophische Studien* 97(1), 63-85.
Del Pinal, Guillermo. “Meaning, Modulation, and Context: A Multidimensional Semantics for Truth-Conditional Pragmatics.” *Linguistics and Philosophy* 41(2), 165–207.
DiFranco, Ralph 2015. “Do Racists Speak Truly? On the Truth-Conditional Content of Slurs.” *Thought: A Journal of Philosophy* 4(1), 28-37.
Dummett, Michael 1981. *The Interpretation of Frege’s Philosophy*. Cambridge, MA: Harvard University Press.
Falbo, Arianna 2021. “Slurs, Neutral Counterparts, and What you Could Have Said.” *Analytic Philosophy* 62(4), 359-375.
Finkbeiner, Rita, Wiese, Heike, and Meibauer, Jörg 2016. “What Is Pejoration, and How Can it be Expressed in Language?” In: *Pejoration*. Edited by Rita Finkbeiner, Jörg Meibauer and Heike Wiese, Amsterdam: John Benjamins Publishing, 1-18.
Fodor, Jerry 1998. *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press.
Frege, Gottlob 1892/1997. “On *Sinn* and *Bedeutung*.” In: *The Frege Reader*. Edited by Michael Beaney, New York: Oxford: Basil Blackwell, 151-171.
Glanzberg, Michael 2011. “Meaning, Concepts, and the Lexicon.” *Croatian Journal of Philosophy* 11(31), 3–31.
Glanzberg, Michael 2014. “Explanation and Partiality in Semantic Theory.” In: *Metasemantics: New Essays on the Foundations of Meaning* Edited by Alexis Burgess and Brett Sherman, New York: Oxford University Press, 259–292.
Glanzberg, Michael 2018. “Lexical Meaning, Concepts, and the Metasemantics of Predicates”. In: *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics*. Edited by Derek Ball and Brian Rabern, New York: Oxford University Press, 197–225.
Gotham, Matthew 2017. “Composing Criteria of Individuation in Copredication.” *Journal of Semantics* 34(2), 333–371.
Gotham, Matthew 2022. “Property Inheritance, Deferred Reference and Copredication.” *Journal of Semantics* 39, 87–116.
Grice, Herbert Paul 1975. “Logic and Conversation.” In: *Studies in the Ways of Words*. Edited by Herbert Paul Grice, Cambridge, MA: Harvard University Press, 22-40.
Harris, Daniel W. 2020. “Semantics without Semantic Content.” *Mind & Language* 37(3): 304-328.
Heim, Irene 1992. “Presupposition Projection and the Semantics of Attitude Verbs.” *Journal of Semantics* 9(3), 183-221.
Hess, Leopold 2021. “Slurs and Expressive Commitments.” *Acta Analytica* 36(2), 263-290.
Hinzen, Wolfram 2007. *An Essay on Names and Truth*. New York: Oxford University Press.
Hinzen, Wolfram, Sheehan, Michelle and Reichard, Ulrich 2014. “Intensionality, Grammar, and the Sententialist Hypothesis.” In: *Minimalism and Beyond*. Edited by Peter Kosta, Steven L. Franks, Teodora Radeva-Bork and Lilia Schürcks, Amsterdam: John Benjamins, 315-349.
Hogeweg, Lotte and Vicente, Agustin 2020. “On the Nature of the Lexicon: The Status of Rich Lexical Meaning.” *Journal of Linguistics* 56, 865-91.
Hom, Christopher 2008. “The Semantics of Racial Epithets.” *The Journal of Philosophy* 105(9), 416-440.

Hom, Christopher and May, Robert 2013. “Moral and Semantic Innocence.” *Analytic Philosophy* 54(3), 293-313.
Hornsby, Jennifer 2001. “Meaning and Uselessness: How to Think about Derogatory Words.” *Midwest Studies in Philosophy* 25, 128-141.
Jeshion, Robin 2013. “Expressivism and the Offensiveness of Slurs.” *Philosophical Perspectives* 27(1), 231-259.
Jeshion, Robin 2020. “Pride and Prejudiced: On the Reclamation of Slurs.” *Grazer Philosophische Studien* 97(1), 106-137.
Kaplan, David 1989. “Demonstratives.” In: *Themes from Kaplan*. Edited by Joseph Almog, John Perry and Howard Wettstein, New York: Oxford University Press, 481-564.
Karttunen, Lauri 1974. “Presupposition and Linguistic Context.” *Theoretical Linguistics* 1(1-3), 181-194. URL
Katz, Jerrold J. 1972. *Semantic Theory*. New York: Harper & Row.
Kennedy, Christopher and Stanley, Jason 2009. “On ‘Average.’” *Mind* 118(471), 583-646.
Kennedy, Randall 2002. *Nigger*. New York: Pantheon Books.
King, Jeffrey C. 2018. “W(h)ither Semantics!(?).” *Noûs* 52, 772-795.
Knowlton, Tyler, Pietroski, Paul, Halberda, Justin and Lidz, Jeffrey 2022. “The Mental Representation of Universal Quantifiers.” *Linguistics and Philosophy* 45, 911–941.
Lasersohn, Peter 2007. “Expressives, Perspective and Presupposition.” *Theoretical Linguistics* 33(2), 223-230.
Lepore, Ernie and Stone, Matthew 2018. *Imagination and Convention: Distinguishing Grammar and Inference in Language*, New York: Oxford University Press.
Lewis, Daniel K. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.
Lidz, Jeffrey, Pietroski, Paul, Hunter, Tim and Halberda, Justin 2011. “Interface Transparency and the Psychosemantics of *Most*.” *Natural Language Semantics* 19(3), 227-256.
McCabe, James 2008. “Paddywhacking and Mick-Taking: Of Being on First-Name Terms with the Irish Other.” In: *L’autre.* Edited by Dove-Rumé, Janine, Michel Naumann, and Tri Tran, Tours: Presses Universitaires Francois-Rabelais, 387-406.
McCready, Elin 2010. “Varieties of Conventional Implicature.” *Semantics and Pragmatics* 3(8) 1-57.
Neufeld, Eleonore 2019. “An Essentialist Theory of the Meaning of Slurs.” *Philosophers’ Imprint* 19(35), 1–29.
Pietroski, Paul M. 2010. “Concepts, Meanings, and Truth: First Nature, Second Nature, and Hard Work.” *Mind and Language* 25, 247-278.
Pietroski, Paul M. 2015. “Framing Event Variables.” *Erkenntnis* 80, 31-60.
Pietroski, Paul M. 2018. *Conjoining Meanings: Semantics Without Truth-Values*. New York: Oxford University Press.
Pritchard, Tim 2022. “Proprietary Linguistic Meaning.” *Synthese* 200, 426. https://doi.org/10.1007/s11229-022-03776-x
Poppa-Wyatt, Mihaela and Wyatt, Jeremy L. 2018. “Slurs, Roles and Power.” *Philosophical Studies* 175(11), 2879–2906.
Potts, Christopher 2007. “The Expressive Dimension.” *Theoretical Linguistics* 33(2), 165-198.
Potts, Christopher 2012. “Conventional Implicature and Expressive Content.” In: *Semantics: An International Handbook of Natural Language Meaning*, Vol. 3. Edited by Claudia

Maienborn, Klaus von Heusinger and Paul Portner, Berlin-Boston: De Gruyter Mouton, 2516-2535.
Pustejovsky, James 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press.
Putnam, Hilary 1975. "The Meaning of 'Meaning.'" In *Language, Mind and Knowledge: Minnesota Studies in Philosophy of Science*, Volume 7. Edited by Keith Gunderson, University of Minnesota Press, 131-193.
Quilty-Dunn, Jake 2021. "Polysemy and Thought: Toward a Generative Theory of Concepts." *Mind & Language* 36(1), 158-185.
Quine, Willard V.O. 1951. "Two Dogmas of Empiricism." *Philosophical Review* 60(1), 20-43.
Ritchie, Katherine 2017. "Social Identity, Indexicality, and the Appropriation of Slurs." *Croatian Journal of Philosophy* 17(50), 155-180.
Sauerland, Uli 2007. "Beyond Unpluggability." *Theoretical Linguistics* 33(2), 231-236.
Schlenker, Philippe 2007. "Expressive Presuppositions." *Theoretical Linguistics* 33, 237-245.
Sennet, Adam 2016. "Polysemy." In: *The Oxford Handbook of Topics in Philosophy* (Online Edition, Oxford Academic, 1 April 2014). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199935314.013.32
Sennet, Adam and Copp, David 2015. "What Kind of a Mistake Is It to Use a Slur?" *Philosophical Studies* 172(5), 1079-1104.
Shakur, Tupac, Elliott, David, Brooks, Ronald and Parker, Maceo 1991. "Violent." [Recorded by Shakur, Tupac.] *2Pacalypse Now* [CD]. San Francisco, CA: Interscope Records.
Stainton, Robert 2006. "Meaning and Reference: Chomskian Themes." In: *The Oxford Handbook of Philosophy of Language*. Edited by Ernie Lepore and Barry Smith, New York: Oxford University Press, 913–940.
Stalnaker, Robert C. 1972. "Pragmatics." In: *Semantics of Natural Language*. Edited by Donald Davidson and Gilbert Harman, Dordrecht: Springer Netherlands, 380-397.
Stanley, Jason 2007. *Language in Context: Selected Essays*. New York: Oxford University Press.
Travis, Charles 1989. *The Uses of Sense*. New York: Oxford University Press.
Vallée, Richard 2014. "Slurring and Common Knowledge of Ordinary Language." *Journal of Pragmatics* 61, 78-90.
Vicente, Agustin 2018. "Polysemy and Word Meaning: An Account of Lexical Meaning for Different Kinds of Content Words." *Philosophical Studies* 175(5), 947-968.
Vicente, Agustin 2019. "Chomskyan Arguments Against Truth-Conditional Semantics Based on Variability and Co-Predication." *Erkenntnis* 86, 919-940.
Waismann, Friedrich 1945. "Verifiability." *Aristotelian Society Supplementary Volume* 19, 101-164.
Wellwood, Alexis, Hespos, Susan J., and Rips, Lance 2018. "The Object : Substance :: Event : Process Analogy." In: *Oxford Studies in Experimental Philosophy*, Vol. 2. Edited by Tania Lombrozo, Joshua Knobe, and Shaun Nichols, New York: Oxford University Press, 183-212.
Whiting, Daniel J. 2008. "Conservatives and Racists: Inferential Role Semantics and Pejoratives." *Philosophia* 36, 375–388.
Whiting, Daniel J. 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, Timothy 2003. "Understanding and Inference." *Aristotelian Society Supplemental Volume* 77, 249-93.

Williamson, Timothy 2009. “Reference, Inference and the Semantics of Pejoratives.” In: *The Philosophy of David Kaplan*, Edited by Joseph Almog and Paolo Leonardi, New York: Oxford University Press, 137–158.
Williamson, Timothy 2010. “The Use of Pejoratives.” In: *The Later Wittgenstein on Language*. Edited by Daniel Whiting, Basingstoke: Palgrave, 45-62.
Zeman, Dan 2022. “A Rich-Lexicon Theory of Slurs and their Uses.” *Inquiry* 65(7), 942-966.
Zwicky, Arnold M., and Sadock, Jerrold M. 1975. “Ambiguity Tests and How to Fail Them.” In *Syntax and Semantics*, Volume 4. Edited by John P. Kimball, Boston-Leiden: Brill, 1-36.