

# **The Sound of Slurs: Bad Sounds for Bad Words**

Eric Mandelbaum (CUNY: Baruch College & the Graduate Center)

Steven Young (Meta)

To appear in *Oxford Studies in Experimental Philosophy*

## Abstract

An analysis of a valenced corpus of English words revealed that words that rhyme with slurs are rated more poorly than their synonyms. What at first might seem like a bizarre coincidence turns out to be a robust feature of slurs, one arising from their phonetic structure. We report novel data on phonaesthetic preferences, showing that a particular class of phonemes are both particularly disliked, and overrepresented in slurs. We argue that phonaesthetic associations have been an overlooked source of some of the more peculiar, arational aspects of slurs. We conclude by drawing broader morals about the evolution of the lexicon.

Consider trash. Pretty much no one likes it. And what holds for the signified holds for the signifier: when asked to rate how much they like a word on a 1-9 scale, with 1 being the lowest rating a participant can give, the word ‘trash’ rates as a 2.74, equivalent to ‘queasy,’ ‘malpractice,’ ‘diabetes,’ ‘tuberculosis,’ and ‘heroin’ (Warriner et al. 2013).<sup>1</sup> ‘Heroin’ has a much greater standard deviation than ‘trash’ (2.6 vs 1.41), so at least some people like ‘heroin,’ whereas pretty much everyone is in agreement that ‘trash’ stinks. It’s even more poorly rated than ‘bigot’ (2.75) and ‘scurvy’ (2.76). Unsurprisingly, synonyms of ‘trash’ also rate poorly: ‘garbage’ rates as a 2.88, equivalent to ‘hyperventilating’ and a bit more disliked than ‘addicted’ (2.89). ‘Litter’ rates as a 2.58, which is a tad worse than ‘colonoscopy’ (2.6), though slightly better than ‘hypothermia’ and ‘bummer’ (both rate at 2.57).

Slurs also tend to be poorly rated words. ‘N\*gg\*r’ rates as a 1.67; ‘w\*tb\*ck’ as 2.75. Surprisingly, words that rhyme with slurs are rated more poorly than their synonyms. ‘Nag’ (2.3; rhymes with ‘f\*g’) is more poorly rated than ‘scold’ (2.77), ‘bother’ (4.44), ‘annoy’ (2.49), and ‘complain’ (3.1). ‘Blink’ (5.25; rhymes with ‘ch\*nk’) is worse than ‘flash’ (5.53) and ‘twinkle’ (6.26). ‘Boon’ (4.84; rhymes with ‘c\*\*n’) is more poorly ranked than ‘benefit’ (6.87), ‘blessing’ (7.18), and ‘bonus’ (8.05). The effect holds even when we move to extreme abstraction: ‘figure’ (5.09) is ranked lower than ‘shape’ (5.41), and ‘number’ (5.59).

What at first might seem like an odd party trick is not some simple trifle. It’s a robust, strange, and hitherto uncovered fact about slurs. But it also is one that holds promise for explaining part of the functional role of slurs. In this paper we aim to elucidate this effect by introducing a view

---

<sup>1</sup> Subjects (n=1827) were all English-speaking American residents who were asked to work quickly and use their first impressions to rate each word, for ~14,000 English lemmas. The scale poles corresponded to how happy (/pleased, satisfied, contented, hopeful) the word made them feel on one end, and unhappy (annoyed, unsatisfied, melancholic, or despaired) on the other.

– phonaestheticism – that is capable of accounting for the rhyming effect along with a range of slurs’ explananda.

Our approach focuses on the phonaesthetic properties of the mental representation of slurs without appealing to their semantics or pragmatics. Most previous theories have proposed that slurs both categorize and derogate and much of the focus has been on analyzing the derogation.<sup>2</sup> Theorists have posited various ways of analyzing derogation: as a semantic content (Hom 2008); as a semantic content plus presuppositional content (Schlenker 2007, Cepollaro 2015); as a semantic content plus a conventional implicature (Williamson 2009, Potts 2012); as a conversational implicature (Nunberg 2018, Bolinger 2015); as a perspective (Camp 2013); and as an expressive meaning (Richard 2008, Potts 2012, Jeshion 2013).

In contrast, we think discussions of slurs have missed an important factor of how slurs function in cognition. What we intend to show is that these two ingredients—categorizing and derogating—aren’t enough on their own to explain people’s reaction to slurs. We will also need a third factor—one that can explain how a slur transmits, in an arational fashion, negative valence to others.<sup>3</sup> To explain this, one needs to appreciate the phonaesthetic value of slurs, which can explain a surprising range deal of slurs’ functional role.

---

<sup>2</sup> Exceptions to this rule include Anderson & Lepore (2013) and Lepore & Stone (2018). Lepore & Stone characterize their theory as including “a catchall description of interpretive effects that go beyond meaning in language,” (2) including “expectations, connotations, associations, analogies and more” (6). We see phonestheticism as a theory offered in a similar spirit.

<sup>3</sup> What is it to transmit in an ‘arational’ fashion? Arational transitions in thought are characterized against rational and irrational transitions. Rational transitions are truth preserving formal inferences corresponding to rationally acceptable rule (e.g., *modus ponens*), or are abductive inferences that conform to normatively acceptable inference rules (Quilty-Dunn & Mandelbaum 2018). Irrational transitions are mistaken applications of one of these rules (e.g., the thinker is trying to run a disjunctive syllogism but infers the wrong disjunct). Arational transitions are not rule based at all—that is, they are not in the purview of intentional or computational psychology, such as associative transitions (Mandelbaum 2016). So, e.g., if one eats oysters and then automatically thinks of Fermat’s

We begin by explicating the backbones of the theory before turning to the data that it can explain.

## 1. The Non-Arbitrariness of Language

A venerable Sausserian thesis is that language is, in a deep sense, arbitrary. The word ‘bird’ seems to have been arbitrarily chosen to refer to birds. In a counterfactual linguistic history ‘bird’ could’ve referred to dogs, or couches, or democracy. The essence of the Sausserian claim is that the “linguistic sign is arbitrary” (1916, 67); any sign could’ve been used to signify any referent.

Although the arbitrariness of language seems broadly correct, there are many counterexamples due to iconicities. For example, although the connection between ‘bird’ and birds seems perfectly arbitrary, the connection between ‘warbler’ and warblers seems less so. Warblers’ songs sound like the name of the birds that sing them; that is, warblers *warble*. ‘Warble’ is a piece of iconic language.

Although defining iconicity is difficult, there are paradigmatic examples. Onomatopoeia is the easiest to conceptualize as it’s an isomorphic mapping of sound to meaning. ‘Fizz’ sounds like *fizzing* but most of language doesn’t work this way. Things get more interesting as we slowly step away from the buzzing, hissing, fizzing onomatopoeics and towards the jiggly, wiggly, squishy, slurpy, bubbly, bloopy, cushy iconicities. Maybe ‘slurp’ sounds like slurping—to slurp is to make a

---

Last Theorem, that associative or (or perhaps brute causal) connection would be arational. For slurs, a rational transition may be one like conversational implicature theory (e.g., Lycan’s theory (2015) where one infers a speaker’s racist attitude from the fact that they’ve uttered a slur in anger). An example of an irrational transition would be someone making a poor inference from the same data (e.g., inferring that someone likes Jews from them uttering ‘k\*ke’ in anger). An arational transition would be a slurs ability to offend regardless of use/mention—that is, regardless of calculating a speaker’s intention (e.g., someone feeling offense from seeing the word ‘k\*ke’ spelled out looking at specimen through a microscope).

sound after all. But ‘jiggly’ both sounds jiggly while meaning more than “to jiggle” (“to move backwards and forward, or up and down, with a light unsteady motion”).<sup>4</sup>

Classic iconicities include the systematic Bouba/Kiki effects (Sapir 1929; Köhler 1947). The rounded ‘Bouba’ gets reliably matched to round faces, while the sharp ‘Kiki’ gets mapped to angular ones so that Renee Zellweger, and Oprah Winfrey, are real Boubas while Audrey Hepburn, and Tyra Banks, are deemed Kikis (Barton and Halberstadt 2018). Similar effects hold for shapes (with ‘Bouba’ applying to dulled edges as in two-dimensional sketches of clouds and ‘Kiki’ to sharp ones as in two-dimensional sketches of stars; Ramachandran & Hubbard 2001; D’onofrio 2014), names (‘Molly’ and ‘Leo’ are Boubas, ‘Katia’ and ‘Tate’ are Kikis, Sidhu & Pexman 2015), and even tastes (sweet being ‘Bouba,’ bitter being ‘Kiki,’ Fairhurst et al. 2015). Bouba/kiki effects are developmentally robust, appearing at 2.5 years at the latest (Maurer et al., 2006).

One can grasp the intuition of sound/referent connections by considering individual cases: how would it feel if we called ants “huge” and hippopotamuses “teeny”? Or compare a belly laugh to a polite giggle; it seems like “guffaws” could never be a snicker, nor ‘tee-hees’ an enormous cackle. But we needn’t just use intuition. Evidence that the ideophonic nature of words is due to sound symbolism and not, say, associative systematicity (such as ‘gl’ signifying shininess in English—‘glint,’ ‘glimmer,’ ‘glitter,’ ‘glow,’ etc.) comes from participants abilities to guess word meanings (using a word bank) in languages they don’t know (Perlman et al. 2015; D’Anselmo et al. 2019).

The foregoing suggests that language is far less arbitrary than Saussure thought. Where language stops being arbitrary it starts becoming systematic in ways that are interesting both for understanding language and, for our purposes, the negativity of slurs. Phonaesthetic preferences run

---

<sup>4</sup> All definitions are from the Oxford English Dictionary (2020).

surprisingly deep. For example, in Aryani et al. (2018), words with short vowels, voiceless consonants, and hissing sibilants (as in ‘piss’) were deemed more arousing and negative than their synonyms (e.g., ‘pee’). Our aim is to extend certain phonetic/valence mappings to the account of slurs. Slurs share an unlikely property: they tend to contain a particular phonetic structure—velar plosives—that is disliked. This sound triggers negative affect, and the negative affect produced by the phonaesthetics is part of the overall negativity experienced by slurs.

## 2. The Phonaesthetics of Slurs and Velar Plosives

Velar plosives--both the voiceless velar plosive ⟨k⟩ and the voiced ⟨g⟩-- are overrepresented in slurs. One finds them in ‘k\*ke,’ ‘d\*ke,’ ‘n\*gger,’ ‘ch\*nk,’ ‘sp\*c,’ ‘w\*tb\*ck,’ ‘cr\*cker,’ ‘kra\*t,’ ‘co\*n,’ ‘mo\*k,’ ‘f\*g,’ ‘f\*ggot.’ Why are these sounds overrepresented in slurs? For one reason, the sounds themselves appear to be intrinsically disliked—they appear at the lowest end of articulation hierarchies, which means they are the phonemes most apt to be avoided (de Lacy 2006; Berent 2017). This is surprising as velar plosives aren’t particularly hard to pronounce—they are acquired relatively early in development and are ubiquitous in world languages (McLeod, S., & Crowe, K 2018).

We begin by investigating whether velar plosives are disliked. We test this prediction by presenting made-up words either containing or lacking velar plosives and asking participants to freely pair each word with either a positive or negative definition (e.g., “The most popular dish in a restaurant” or “An illness caused by inhaling chemicals”).<sup>5</sup> Novel non-words containing velar plosives (‘leckin’) were found to be liked less than their matched equivalents (‘leshin’). Participants

---

<sup>5</sup> The made-up pseudowords were possible but non-actual English words (i.e., they were all pronounceable), and are listed in the SI as experiment 1, as well as the full design, procedure, and results.

were 46% more likely to match the words containing the velar plosive ('ukib') to the negative definition and the matched control ('unib') to the positive one, a robust ( $d = .65$ ) and statistically significant difference ( $p = .03$ ).

To ensure the robustness of the result we replicated and extended the results in an even simpler task. Instead of asking subjects to match the novel words to definitions, here participants were just asked to guess whether the word meant something positive or negative. Participants were more 71% apt to match words with a velar plosive to negativity than positivity ( $p = .02$ ).<sup>6</sup>

To better understand slurs' power to affect other words through mere phonetic similarity, we analyzed a corpus of word valences, which lead to the observations at the start of this essay (details listed as experiment 3 in the SI). If phonaestheticism accounts for some of the derogatory force of slurs, and in doing so explains how slurs differ from their default counterparts, then we should expect that words that are associated with slurs by mere phonetic similarity (i.e., containing velar plosives), should be contaminated by the negative valences connected to those slurs. After reviewing 107 words that rhyme with slurs and over 370 synonyms for those words, we found that, as predicted by phonaestheticism, people disliked words that rhyme with slurs (e.g., 'vigor') more than their synonyms ('effort' [6.47], 'strength' [6.73], 'health' [6.85], and 'energy' [6.9]) 62% of the time, well above chance both descriptively and statistically ( $p < .05$ ). The reason why this held was that words that rhyme with slurs themselves are more likely to contain a velar plosive. It's the velar plosives—and not, e.g., a word's overall associative connections<sup>7</sup>—that are causing words to be more disliked than their synonyms.

---

<sup>6</sup> Full design, procedure, and results are in the SI (experiment 2).

<sup>7</sup> If mere associative connections were explaining the effect, then we'd expect the most strongly negatively valenced words that don't contain velar plosives (e.g., 'divorce') to also display the effect. Yet words that rhyme with the most strongly negatively valenced words ('endorse') aren't, on

Most straightforwardly, following Warriner et al. (2013), we asked subjects to report how happy or unhappy, excited or calm, and in control or out of control each of our made-up words made them feel.<sup>8</sup> Answers were averaged to form a reliable composite ( $\alpha > .77$ ) measure of word liking. We coded responses such that higher scores indicated greater liking. This found that words with the presence of a velar plosive ( $M = 4.80$ ,  $SD = .74$ ) were liked less than control words ( $M = 4.98$ ,  $SD = .57$ ),  $p = .006$ ,  $d = .36$ , even when the “words” were artificial and had no prior associations.

We can be reasonably confident that velar plosives are disliked, and also overrepresented in slurs. Perhaps their historical placement in slurs makes them more apt to be disliked, or perhaps because of primitive phonaesthetic preferences they are disliked and so were deployed when coining pejoratives. It could be that markedness is doing the work, or it could be that placement on articulation hierarchies is (velars have the lowest placement; that is, they are most likely to be avoided de Lacy 2006). All of these explanations could be simultaneously true. But regardless of why velar plosives are disliked, there being so can go some way into explaining part of the functional role of slurs without averting to pragmatic or semantic accounts. In particular, it can help explain some of the peculiar irrationality of slurs.

### 3. Use/Mention Insensitivity and Phonetic Similarity

---

average, disliked more than their synonyms (‘apply’). See SI section experiment 3—‘low valence non-slur words’—for details.

<sup>8</sup> These scales were defined for participants during a practice round using the same language employed by Warriner et al. Each subject was shown 20 words in random order, 11 containing velar plosives and 9 control words without any velar plosives. Responses for happiness, excitability, and in control questions were reported on a scale of 1 to 9. Ease of pronunciation responses were reported on a scale of 1 to 7. 76 participants were recruited through Mechanical Turk. They were paid twenty cents (\$0.20), 51 were women. This sample was determined by the observation of large effects in Experiment 4 ( $d = .64$ ) and provides high power (.918).

Slurs exert derogatory force when embedded under negation, in belief reports, in quotation, and when otherwise mentioned and not used. Consider the following:

- (1) Adam is not a f\*gg\*t.
- (2) Kim said her neighbor is a kr\*ut.
- (3) The word 'sp\*c' contains four letters.

The slurs in these sentences cause offense despite 1's use of negation, that 2 is a report of someone else's utterance, and that 3 is an instance in which the slur is merely mentioned.

Phonaestheticism can straightforwardly explain these data. Phonaesthetic associations are activated regardless of the context in which the target concept occurs. "Lawyers aren't sharks" activates SHARKS as well as the concepts and mental ephemera associated with sharks (Hasson & Glucksberg 2006). Similarly, if a slur is uttered, the corresponding phonaesthetic association—the sound/negative feeling pairing—is activated regardless of whether it is merely being mentioned or negated. The negative affect then automatically activates associated negative concepts in the normal spreading activation way (Mandelbaum 2020). Negative concepts, attitudes, and valences are activated before a speaker's intent can be inferred. In fact, they are activated even when one may know the speaker holds the opposite intent, or even when the speaker is speaking in another language.

Perhaps most importantly, we know from attribution theory and dissonance theory (as well as the metacognitive fluency literature) that we feel the need to explain the causes of our feelings (Festinger 1957; Schachter & Singer 1962; Oppenheimer 2008). When we are feeling positive or negative, we don't assume that it's merely a chemical change, but instead search for the reason why we are feeling that way; this especially holds when we feel negative (Bohner et al. 1988). A hypothesis that isn't generally salient is that there are certain intrinsic phonaesthetic associations

linking the words we hear (or use) to our affective state. So instead, we confabulate and infer that there must be some ill-will afoot.

For instance, consider the controversy over the USC business school professor Greg Patton, who was teaching a lesson on filler terms (e.g., ‘er,’ ‘um’) in other languages.<sup>9</sup> Professor Patton said that in China the common pause word translated as ‘that’ said over and over again as ‘ne ga, ne ga, ne ga.’ Although an entirely different language was being uttered, the phonetic proximity between ‘ne ga’ and the English slur word for African Americans was enough to cause an extremely negative reaction in some of the hearers.<sup>10</sup> A phonaesthetic account can explain why: it’s the mere association of the sounds that brings an inherent negativity regardless of the semantics or pragmatics of the words (which in this case the students presumably didn’t even know).

The scenario should be somewhat familiar. When bad things happen—i.e., when we feel bad—we search for an explanation of why this is the case (Lerner 1980; Epley et al. 2007; Gray & Wegner 2010; Mandelbaum & Ripley, 2012). So, the causal chain goes from perception of velar plosives (e.g., hearing the sounds), to an experience of negative affect, to a search for the cause of the negative affect which, often enough, is attributed to the speaker’s intentions or the situation at large.

A similar tactic can be used to explain how the derogatory force of slurs is autonomous from the occurrent, consciously endorsed attitudes of slur producers. Camp observes that slur users do not always employ these offensive words as “verbal weapons,” but they also use them in casual, emotionally neutral conversation (Camp 2013). Further, Camp notes that “many slurs appear to

---

<sup>9</sup> <https://www.insidehighered.com/news/2020/09/08/professor-suspended-saying-chinese-word-sounds-english-slur>

<sup>10</sup> This is reminiscent of Lepore discussing David Howard, where he pointed out that incidental uses can “cause as much a stir as standard offensive language” (Lepore 2010).

permit the *denial* of any negative feeling” and may be used to express some explicit attitude of admiration (332).

Thus, we need not only capture how slurs operate as verbal expressions of aggression and contempt, but also how they are used “in relaxed conversation among bigots” (Camp 331). Slurs used in these casual contexts are derogating, nonetheless. Ill-intent is not necessary for a word to operate as a slur.

To see how phonaesthetic associationism deals with examples like these, consider how it severs the intention of the speaker from the offensive value of the utterance. According to phonaestheticism, if a bunch of seashells randomly aligned on the shore to spell out ‘k\*ke,’ that would be offensive to onlookers, even though no speech act occurs. A hearer encountering a pejorative automatically (and arationally) activates its negative network of associations regardless of the source of the pejorative. Thus, we can know that the casual bigot is sincerely denying any negative feeling and yet still feel offense as the mere mention of the slur carries negative affect, and negative affect demands rationalization.

#### **4. What Do We Want a Theory of Slurs for Anyway?**

Even if phonaestheticism is correct would that close the book on an account of slurs? To answer this question, we need an idea of what work we want a theory of slurs to do. There are many different explanatory goals one might have in mind. For one thing, one might want to answer the individuation question—what makes slurs qualitatively different from their neutral counterparts. But one might also want to answer a different question, one about the offensiveness of slurs. Here what we want is a theory of why people behave in a certain way—why they behave in the canonical way people do when encountering slurs. And all behavior—save for maybe extreme cases of reflexes

(e.g., the deep tendon reflex) is always an interaction effect of multiple sources. The most any theory should ever hope to explain is some portion of the variance of behavior. A theory of much simpler behavior—say hand raising—is itself going to be complex, with the behavior only weakly predictable even after accounting for both implicit and explicit attitudes. The cognition of slurs is a much more complex phenomenon. Seen from this vantage point, competing theories aren't exclusive competitors. As we see it, semantic, implicature, and expressivist, and prohibitionist theories are wrongly seen as combatants for the true theory of how slurs work; instead, each may be picking up on different aspects of a complex phenomenon of how slurs derogate.

Our theory adds to this list by offering an overlooked non-semantic and non-pragmatic factor. In that sense it's in the orbit of Anderson and Lepore (2013) and Lepore and Stone (2018)'s 'Prohibitionist' theories. For example, Lepore and Stone describe their view "as a catchall description of interpretive effects that go beyond meaning in language," (2) including "expectations, connotations, associations, analogies and more" (6). Phonaesthicism is working out how associations work—or don't work—by using a more detailed account of the phonetics of slurs.

To see the lacuna our theory fills, take, for example, Nunberg who criticizes Hornsby (2001) by adding italics to her own words: "About derogatory words...one finds oneself saying that negative or hostile attitudes of *their users* have rubbed off onto them" (Nunberg, 43). Nunberg is making a point about conversational implicature, and then inferring the intentions of users. In contrast, we think discussions of slurs have missed an important factor of how slurs work in cognition. To see this factor, all we need to do is to change the italics from *their users* to *rubbed off*. What phonaesthicism shows is that categorizing and derogating aren't enough on their own to explain how slurs affect people—we will also need a third factor—one that can explain how a slur's pejoration just  *rubs off* on others, as if slurs were a verbal virus. One might have thought—as Lepore

and Stone imply—that to understand this one must appreciate the associative value of slurs. Phonaesthicism shows that it's not the associative properties of words on their own but their phonetic properties that are critical to this explanation.<sup>11</sup> Moreover, the phonaesthetic properties of slurs aren't a mere accident—words have been coined and selected for a reason, a topic we return to presently.

## 5. Phonaesthetics and the Euphemism Treadmill

Imagine we had to make up a set of new pejoratives right now. How do people come up with new slurs? Perhaps people unconsciously know that velar plosives are negatively valenced. This implicit knowledge is then codified in the lexicon: as new words for slurs are coined, they are built with negative phonemes to match the negative valence the coiners feel towards the groups' referent. The lexicon thus encodes some information about speaker's attitudes: the phonetic properties of words contain information about how users think of their referents. This is the 'sound symbolism' hypothesis and one can see it in iconic language across the lexicon (Dingemanse et al. 2015). These iconicities are useful; for example, they aid in language acquisition (Perry et al. 2018) and function as communicative lubricants. Effects of phonetics on word coinage abound. For instance, Pokemon names encode information about the characters size, weight, and strength (*inter alia*; Kawahara et al. 2016), and baseball player nicknames carry similar information (Shih & Rudin 2020).

---

<sup>11</sup> Although we've gone further than previous theorists in investigating the phonetic and associative properties of slurs we have not come anywhere near a full explanation of their roles. Take the case of Prof. Patton. What if instead of saying 'ne ga' he uttered 'ga ne'—would that have caused offense? Presumably not. This suggests it's not just the utterance of the phonemes, but their order and how that order triggers items in the mental lexicon. (We thank the editors for making this point to us as they suggested this example. If we missed this point, there are surely scores more subtleties our phonetic account misses. Instead of closing the book on the phonetic aspect of slurs we see our theory as an opening for others to investigate how these properties work.)

If sound symbolism is correct, it suggests something disappointing about coining new terms to stem offense. To see why, consider the question: could changing disliked phonemes substantially change anything about social categorization?

Suppose everyone knows that Donald Trump's favorite word is a particular slur. The New York Times would allow us to use 'Trump's favorite word' to refer to the slur. However, over time 'Trump's favorite word' will itself become entrenched in its association to the slur. And then use of 'Trump's favorite word' would also be prohibited in the same ways that other close associates of that slur are.

This is a losing battle. As has been forever noted, Eskimos may not have 12 words for snow, but English speakers absolutely do have more than 12 words for bathroom (and toilet, some of which metonymically stand for bathroom): 'lavatories,' 'lavs,' 'loo,' 'w.c.,' 'restrooms,' 'facilities,' 'bog,' 'facilities,' 'comfort station,' 'washroom,' 'men's/women's room,' 'john,' 'privy,' 'smallest room in the house,' 'crapper,' 'can,' 'potty,' 'pot,' 'throne,' 'stool,' 'commode,' 'can,' etc. 'Bathroom' was introduced to push 'house-of-office,' 'privy-house,' 'bog-house,' among others out of the lexicon. Note that these are all sanitized from the words that could, in a compositional way, more clearly express the meaning of the place: 'shithole,' 'defecation place,' 'poopland,' etc.

Why have so many words for bathroom categories, but not for, say, bicycle? It's due to 'The Euphemism Treadmill' (Pinker 2002; the above examples are expanded from his list). The Euphemism Treadmill refers to a phenomenon where each generation creates a new, neutral to positively valenced word or phrase (e.g., 'palmetto bug'; 'janitor') to replace a negatively valenced word ('cockroach'; 'custodial engineer'). Over time the negative valence attaches to the new word. The Treadmill occurs because our concept of the word's referent is a strongly negatively valenced conditioning stimulus. We attempt to give a rosier take on the previously negatively valenced word,

trading 'bathroom' for the serene '*restroom*,' but inevitably 'restroom' gets marred with the same valence. Once the term gets entrenched, it takes on the valence of its referent through evaluative conditioning. This process, which has itself gone by many names that are analogous to Euphemism Treadmill (e.g., 'pejoration,' 'semantic change,' and 'evaluative conditioning') is what ensures that we are fighting a losing battle when hoping that restricting usage of a verboten term will cause any lasting change. All that's bound to happen is that the new term will, in time, pick up the same negative connotation the old slur had.

The Euphemism Treadmill thus points us to sobering conclusion: there will be no word so positively valenced, nor any construction so circuitous that it can't pick up the valence of extremely negatively valenced referents. Rather, the solution is to get to a point where we no longer have negatively valenced mental representations of disadvantaged groups (or better, where we no longer have disadvantaged groups). Otherwise, problematic aspects of slurs will continue apace, regardless of how successful certain appropriation or reclamation projects are.<sup>12</sup>

## 6. Conclusion

Contrary to assumptions about the arbitrary connection between references and referent, the present paper argues that the part of the powerful negativity of slurs is derived from systematic properties of the words themselves. Specifically, the overrepresentation of intrinsically negative velar plosives in slurs is a phonaesthetic source of negativity i.e., slurs simply sound bad, and thus these phonetic forms are frequently used in words that refer to disliked and stigmatized groups of people.

---

<sup>12</sup> Linguistic reclamation is the process through which slurs lose their negativity after being retaken by a stigmatized group. For example, the gay community using f\*g as an ingroup signifier. We encourage future research on this topic from a phonaesthetic perspective. However, a deep dive into the various social linguistic, and cognitive mechanisms associated with this process is beyond the scope of this brief report.

The inherent negativity of slurs leads to automatic associative processes upon encountering these charged words, further reestablishing the negative affect induced by slurs; this is so even when the use of slurs is incidental or ostensibly non-hostile (as in the case of the proverbial “casual bigot”). Finally, even euphemistic phrases designed to avoid the impolite and phonaesthetically displeasing qualities of slurs eventually become contaminated, in part through evaluative condition. This renders efforts to repurpose slurs challenging, so long as the groups they refer to as still subject to systematic prejudice. But perhaps by understanding the arational aspects of slurs we will better understand the myriad ways we tend to think about referents society deems distasteful.

## References

- Anderson, L & Lepore, E. (2013). 'What Did You Call Me?' Slurs as Prohibited Words. *Analytic Philosophy* 54 (3): 350-363.
- Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018). Why 'Piss' is Ruder than 'Pee'? The Role of Sound in Affective Meaning Making. *PloS one*, 13(6), e0198430.
- Barton, D. N., & Halberstadt, J. (2018). A Social Bouba/Kiki effect: A Bias for People whose Names Match their Faces. *Psychonomic Bulletin & Review*, 25(3), 1013-1020.
- Bohner, G., Bless, H., Schwarz, N., & Strack, F. (1988). What Triggers Causal Attributions? The Impact of Valence and Subjective Probability. *European Journal of Social Psychology*, 18(4), 335-345.
- Berent, I. (2017). Is Markedness a Confused Concept?. *Cognitive Neuropsychology*, 34(7-8), 493-499.
- Camp, E. (2013). Slurring Perspectives. *Analytic Philosophy* 54 (3): 330-349.
- Cepollaro, B. (2015). In Defense of a Presuppositional Account of Slurs. *Language Sciences* 52: 36-45.
- D'Anselmo, A., Prete, G., Zdybek, P., Tommasi, L., & Brancucci, A. (2019). Guessing meaning from Word Sounds of Unfamiliar Languages: A Cross-Cultural Sound Symbolism Study. *Frontiers in Psychology*, 10, 593.
- D'Onofrio, A. (2014). Phonetic detail and dimensionality in sound-shape correspondences: Refining the bouba-kiki paradigm. *Language and speech*, 57(3), 367-393.
- De Lacy, P. (2006). *Markedness: Reduction and Preservation in Phonology* (Vol. 112). Cambridge University Press.
- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences*, 19(10), 603-615.
- Epley, N., A. Waytz, and J. Cacioppo. (2007). On Seeing Human: A Three Factor Theory of anthropomorphization. *Psychological Review* 114(4): 864-886.
- Fairhurst, M. T., Pritchard, D., Ospina, D., & Deroy, O. (2015). Bouba-Kiki in the Plate: Combining Crossmodal Correspondences to Change Flavour Experience. *Flavour*, 4(1), 22.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Palo Alto: Stanford University Press.
- Gray, K., & Wegner, D. M. (2010). Blaming God for Our Pain: Human Suffering and the Divine Mind. *Personality and Social Psychology Review*, 14(1), 7-16.
- Hasson, U., & Glucksberg, S. (2006). Does Understanding Negation Entail Affirmation?: An Examination of Negated Metaphors. *Journal of Pragmatics*, 38(7), 1015-1032.
- Hom, C. (2008). The Semantic of Racial Epithets. *Journal of Philosophy* 105 (8): 416-440.
- Hornsby, J. (2001). Meaning and Uselessness: How to Think About Derogatory Words. *Midwest Studies in Philosophy* 25.1 (2001): 128-141.
- Jeshion, H. (2013). Expressivism and the Offensiveness of Slurs. *Philosophical Perspectives* 27 (1): 231-259.
- Kawahara, S., Noto, A., & Kumagai, G. (2016). Sound (symbolic) patterns in Pokémon names: Focusing on voiced obstruents and mora counts. *PLoS ONE*, 1-14.
- Köhler, W. (1947). *Gestalt psychology* (2nd edn.). New York: Liveright Publishing
- Lerner, M. J. (1980). The Belief in a Just World. In *The Belief in a just World* (pp. 9-30). Springer, Boston, MA.

- Lepore, E. (2010). Speech and Harm. Retrieved January 26, 2016.  
 <<http://opinionator.blogs.nytimes.com/2010/11/07/speech-and-harm/>>.
- Lepore, E. & Stone, M. (2018). Pejorative Tone. In D. Sosa (ed) *Bad Words*. Oxford: Oxford University Press.
- Lycan, W. G. (2015). Slurs and lexical presumption. *Language Sciences*, 52, 3-11.
- Mandelbaum, E., & Ripley, D. (2012). Explaining the abstract/concrete paradoxes in moral psychology: The NBAR hypothesis. *Review of Philosophy and Psychology*, 3(3), 351-368.
- Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, 50(3), 629-658.
- Mandelbaum, E. (2020). Associationist theories of thought. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), URL = <https://stanford.library.sydney.edu.au/archives/spr2015/entries/associationist-thought/>
- Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The Shape of Boubas: Sound–Shape Correspondences in Toddlers and Adults. *Developmental Science*, 9(3), 316-322.
- McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American journal of speech-language pathology*, 27(4), 1546-1571.
- Nunberg, G. (2018). The Social Life of Slurs. ID. Fogal, D. Harris, and M. Moss (eds.) *New Work on Speech Acts*. New York: OUP.
- Oppenheimer, D. M. (2008). The Secret life of Fluency. *Trends in Cognitive Sciences*, 12(6), 237-241.
- Oxford English Dictionary Online. September 2020. Oxford University Press.
- Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity Can Ground the Creation of Vocal Symbols. *Royal Society Open Science*, 2(8), 150152.
- Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science*, 21(3), e12572.
- Pinker, S. (2002) *The Blank Slate: The Modern Denial of Human Nature*. New York: Penguin Books.
- Potts, C. (2012). Conventional Implicature and Expressive Content. C Maienborn, K. von Heusinger, & P. Portner (eds.) *Semantics: An International Handbook of Natural Language Meaning, Volume 3*. 2516-2536. Berlin: Mouton de Gruyter.
- Quilty-Dunn, J., & Mandelbaum, E. (2018). Inferential transitions. *Australasian Journal of Philosophy*, 96(3), 532-547.
- Ramachandran, V.S., & Hubbard, E.M. (2001). Synaesthesia – A Window into Perception, Thought, and Language. *Journal of Consciousness Studies*, 8 (12), 3–34.
- Richard, M. (2008). *When Truth Gives Out*. New York: Oxford University Press.
- Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5), 379.
- Sapir, E. (1929). A Study in Phonetic Symbolism. *Journal of experimental psychology*, 12(3), 225.
- Saussure, F. D. (1919). (1972). *Course in General Linguistics* (R. Harris, Trans.). La Salle, IL: Open Court.
- Schlenker, P. (2007). Expressive Presuppositions. *Theoretical Linguistics* 33(2): 237-245.
- Sidhu, D. M., & Pexman, P. M. (2015). What's in a name? Sound symbolism and gender in first names. *PloS one*, 10(5), e0126809.
- Shih, S. S., & Rudin, D. (2020). On sound symbolism in baseball player names. *Names*, 1-18.
- Warriner, A., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods* 45 (4): 1191-1207.

Williamson, T. (2009). Reference, Inference, and the Semantics of Pejoratives. J. Almog & P. Leonardi (eds.) The Philosophy of David Kaplan.