University of Kent

Kent Academic Repository

Landes, Ethan (2025) *Is it a tastytaste or a greedgrab? The importance of label choice in language design*. Philosophical Psychology, 39 (8). pp. 3157-3184. ISSN 0951-5089.

Downloaded from
https://kar.kent.ac.uk/117474/ The University of Kent's Academic Repository KAR

The version of record is available from
https://doi.org/10.1080/09515089.2025.2531401

This document version
Publisher pdf

DOI for this version

Licence for this version
CC BY (Attribution)

Additional information

Versions of research works

Versions of Record
If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts
If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in ***Title of Journal*** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries
If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).



Philosophical Psychology

ISSN: 0951-5089 (Print) 1465-394X (Online) Journal homepage: www.tandfonline.com/journals/cphp20

Is it a tastytaste or a greedgrab? The importance of label choice in language design

Ethan Landes

To cite this article: Ethan Landes (2026) Is it a tastytaste or a greedgrab? The importance of label choice in language design, Philosophical Psychology, 39:8, 3157-3184, DOI: 10.1080/09515089.2025.2531401

To link to this article: https://doi.org/10.1080/09515089.2025.2531401

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 17 Jul 2025.

Submit your article to this journal

Article views: 654

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=cphp20

PHILOSOPHICAL PSYCHOLOGY
2026, VOL. 39, NO. 8, 3157–3184
https://doi.org/10.1080/09515089.2025.2531401

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Is it a tastytaste or a greedgrab? The importance of label choice in language design

Ethan Landes

Department of Psychology, University of Kent, Canterbury, UK

ABSTRACT

Everyone who has come up with a new idea or concept faces the same problem: they need to call it *something*. This choice is simultaneously arbitrary and important. The new idea could have any name, but some names for it are clearly better than others. This paper builds upon existing research on the perceived relationship between label and word meaning to better understand the way that label choice affects language design, particularly conceptual engineering. Across three experiments, the paper finds that when learning novel concepts, the labels serving as vehicles for the novel concepts affect how people perceive the novel concepts' content. In particular, the experiments find that labels influence people's initial perception of novel concepts' valence. Not all forms of information are readily extracted from labels, however, and subtle forms of information, particularly phonesthemes, do not appear to influence how people learn new concepts. Accordingly, conceptual engineers and other forms of language designers need to do careful research into how a given label will affect how their novel concept is interpreted by others.

ARTICLE HISTORY
Received 13 August 2024
Accepted 11 June 2025

KEYWORDS
Conceptual engineering; language design; experimental philosophy; philosophy of language; conceptual propagation

1. Introduction[1]

On the one hand, it is a near truism – often credited to Saussure de, F. et al. (1916) – that the relationship between words and meaning is arbitrary. Any series of sounds can have any meaning. On the other hand, we intuitively know that some names are better than others. This is a familiar phenomenon for any academic who has presented their research. You have come up with a new distinction or found some new phenomena that you want to include in a paper. But the name for it is elusive. Nothing quite *feels* right. Presenting the research to colleagues, you give it the best name you can, and despite your best efforts, one of the colleagues tells you that the name you were using is distracting and bad. Then, months later, the *perfect* label comes to mind – it conveys exactly the right idea of what it is you are talking about.

CONTACT Ethan Landes landese@outlook.com University of Kent, Canterbury, UK

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

But how do we make sense of and measure the goodness or badness of a label?

The question is of particular interest to anyone engaged in linguistic activism – the deliberate use of specific words or phrases for ethical or political ends. One form of linguistic activism, conceptual engineering, is the practice of identifying problems with our existing concepts and then engineering revisions or replacements (Isaac et al., 2022). Conceptual engineering is purpose-focused, aiming to make some change in the world by introducing or changing concepts. The concept cannot float freely, however, and so conceptual engineers need some label to serve as the vehicle for this concept. There has been considerable debate over what makes one label or another appropriate and in what cases, for example, it would be appropriate to reuse an old label for new ends (Cappelen, 2018; Chalmers, 2020; Haslanger, 2000). These cases have, however, largely proceeded based on speculation about how people would interpret the new label, although a few authors have used qualitative research or past quantitative research to make more informed empirical posits about the importance of labels as part of conceptual engineering (Fischer, 2020; Koch, 2024; Koslow, 2022; Landes, 2025a; Neufeld, in press).

While conceptual engineering is a method under debate within philosophy, this literature should be understood as clarifying and building upon a preexisting and widespread practice. Indeed, considerable effort has been spent by activists of various stripes trying to change or introduce labels because of the sense that labels themselves are in some way important. For example, feminist linguistic activism (Cooper, 1990; Milles, 2011a) boasts an impressive array of successes in changing language and introducing novel terms (often as vehicles for novel concepts). This includes the successful introduction of "sexual harassment", "cisgender", and "transgender", the replacement of the generic "he" in contexts where the gender is not known with the singular "they", and the ongoing campaigns to explicate or clarify the distinction in meaning between "sex" and "gender".

While the literature of conceptual engineering typically focuses on examples of linguistic activism in English, linguistic activism is a global phenomenon. For example, Swedish feminist linguistic activists boast a unique set of successes due to unique features of Sweden's political environment (Milles, 2011a, 2011b). Perhaps the most notable success in Swedish arose out of the recognition that while young males had child-friendly language to describe their genitalia ("snopp", which is typically translated to English as "willy"), young females did not. This is a feature English shares, but unlike in English, in the early 2000s feminist campaigns in Sweden successfully introduced "snippa" into Swedish as a playful and catchy female counterpart to the male "snopp" (Milles, 2011b). This is not the only example unique to Swedish. Feminist linguistic activists have also pushed Swedish speakers, albeit with

less success, to refer to the hymen in Swedish with the new label "slidkrans" ("vaginal corona") instead of the established "mödomshinna" ("virginity membrane") (RFSU, 2017). Activists argue that not only is "slidkrans" more anatomically accurate than "mödomshinna" but that the shift in language can help highlight and undermine patriarchal assumptions about female sexuality and purity (Milles et al., 2018).

"Snippa" and "slidkrans" illustrate two distinct intuitions about the effect of labels on people's perceptions, what we might call *addition effects* and *frame effects*. First, "Snippa" was introduced to fix a gap in the expressive power of Swedish that was seen to cause real-world harm. "Snippa" was introduced to have effects of addition. The intuition here is that the very existence of a label has a causal effect on language users, and something about the existence of a label, especially a good label, allows people to better grasp and talk about the phenomena (see Fricker, 2007; Maitra, 2018). Empirical work bears out intuitions about the effects of addition. For example, in tasks where participants have to learn novel categories, giving those novel categories labels help people learn the novel categories (Lupyan & Casasanto, 2015; Lupyan et al., 2007).

In contrast to "snippa", "slidkrans" ("vaginal corona") is seen as a *better* label for the hymen than "mödomshinna" ("virginity membrane"). The goal here was not an addition effect meant to increase Swedish speakers' expressive power – the two terms are taken to have identical referents. Instead, activists aim to frame the referent differently, as "slidkrans" avoids the unwanted connotations and imagery of "mödomshinna" (Milles et al., 2018). Accordingly, the goal of relabeling is typically more than merely changed linguistic beliefs about what words mean, aiming at changing non-linguistic beliefs and behaviors (Landes, 2025a, 2025b). As Swedish campaigners argue, the reference to virginity in "mödomshinna" helps support patriarchal attitudes about female sexuality while causing false beliefs about female anatomy, and "slidkrans" aims at addressing both (Zaltzman, 2023).

Notice that frame effects are not just goals of relabeling campaigns. Labels for otherwise unlabeled things can also be better or worse relative to some goal, and many terms that potentially cause addition effects may also cause frame effects. For example, while "snippa" gave a name to something without one, the symmetry with the male "snopp" framed the corresponding genitalia as two sides of a sexual binary (Milles, 2011b, p. 102). The joint desiderata of addition effects and frame effects explains part of the reason why coming up with a name for a new thing is so hard. It is not enough that we give it *some* name, we need to give it the *right* name.

2. The relationships between label and meaning

For the attempts to harness the power of labels with frame effects to pay off in the way linguistic activists hope, different labels do in fact have to

influence language users differently. Evidence of whether frame effects can achieve such wide-ranging change is, however, currently only circumstantial. The most abundant evidence of this sort of causal relationship comes from the countless patterns of regularity between form and meaning across human natural language. At the most basic level, there are statistical regularities, called *systematicity*, between form and meaning (Dingemanse et al., 2015). For example, English verbs are, on average, shorter than English nouns and have proportionately fewer vowels (Cassidy & Kelly, 1991, 2001; Monaghan et al., 2007). While systematicity can be principled or motivated, the notion does not appeal to such connections. In contrast, *iconicity* is when the form of language resembles its meaning. The most obvious forms of iconicity include onomatopoeias like "moo" which sound like their referent or sign language signs that look like the action they represent (Baus et al., 2013; Grote & Linz, 2003).

These patterns affect people's representations of language. For example, funny words are more likely to be iconic and have more surprising and complex sound combinations than other English words (Dingemanse & Thompson, 2020; Westbury & Hollis, 2019). More fundamentally to the relevance to language design, *sound symbolism* – the resemblance between *sound* and meaning (Aveyard, 2012; Klink, 2000; Nuckolls, 1999; Sidhu & Pexman, 2018) – has been found to affect language learning. Japanese 3-year-olds learn sound-symbolic words better than non-sound-symbolic verbs (Imai et al., 2008). Studies on infants have found infants react differently to sound symbolically concordant pairings of shapes and sounds vs discordant pairings (e.g., calling a spikey shape "kiki" vs "moma") in ways that suggest the discordant pairings are more surprising and more difficult to process (Asano et al., 2015; Ozturk et al., 2013; Peña et al., 2011). This is not just limited to sound symbolism, as systematic words are more common among the English words typically learned at early ages than English words typically learned later in life (Monaghan et al., 2014).

In adults, sound symbolism has long been known to influence what sorts of labels people naturally pair to objects of definitions (Nuckolls, 1999; Sapir, 1929). Recently, Mandelbaum et al. (2024) examined the role of phonesthemes in English slurs. Phonesthemes are the systemic pairing of sounds with meaning. Mandelbaum et al. tested how velar plosives (k,g), a phonestheme, affect people's guesses about word meaning. Creating pairs of pseudowords with and without velar plosives ("vlack" vs "vlarp"), Mandelbaum et al. found participants were more likely to guess pseudowords with velar plosives had negative definitions than their counterparts. This was true in tasks where specific good and bad definitions were offered as options ("An illness caused by inhaling chemicals" or "a celebration of cultural heritage") and in tasks where participants were asked to guess, in the abstract, whether the word had a positive or negative definition. This

suggests that subtle sound-symbolic information below the level of morphemes influences interpretation.

Related research exists in marketing. Marketing firms introducing new brands face similar challenges as language designers regarding labeling. Like with new concepts or definitions, new brands need some vehicle through which to communicate their content. A company cannot merely sell carbonated sugar water with artificial caramel flavoring; a company needs to sell Coca-Cola. Research on brand names has found that the names themselves have effects on perceptions of the brands themselves. Female-coded names, such as Nestlé, lead to warmer associations with brands, which positively affects trust and brand choice (Pogacar et al., 2021; see also Wu et al., 2013). Brands with repetitive names, such as Lululemon, similarly increase positive effect (Argo et al., 2010).

However, it is not clear how robustly the information contained in labels affects people's interpretations of word meaning or reference when they encounter them for the first time. One study manipulating the name of a fictional country in a vignette about the acceptability of preemptive strikes found no effect of fine-grained lexical information like sound symbolism (Majnemer & Meibauer, 2023). Majnemer and Meibauer examined the differences between the judged acceptability of preemptively striking "Celesta", "Minalo", and "Drakhar" – fictional country names designed to contain either familiar or unfamiliar letters and sound combinations as well as submorphemic information meant to encourage positive or negative value judgments. "Celesta" was designed to be the most positive and familiar while "Drakhar" was designed to be the most negative and unfamiliar, but scores between the two countries were nearly identical. The study did find lower acceptability scores for "Kerejistan" than the other fictional countries, suggesting more coarse-grained morphemic information, specifically the suffix "–stan", affected judgments in a way submorphemic information did not.

The goal of this paper is to build on existing empirical research to explore and clarify the effect of labels on people's initial interpretation of novel ideas. Three experiments explore a type of language design particularly relevant to conceptual engineering and other forms of linguistic activism, namely, instances where labels are used as vehicles for novel concepts or ideas. By better understanding the way a label influences the perception of novel semantic or conceptual information, conceptual engineers and other linguistic activists can better design their labels to reach their desired ends.

To operationalize this, the studies below examine whether the perceived valence of novel concepts can be affected by positively or negatively loaded labels. Do labels influence how good or bad the content of a new concept is perceived? Experiments 1 and 2 focus on information at the level of morphemes. Using novel concepts and labels that contain

recognizable morphemes, Experiment 1 tests whether labels composed of familiar words can frame novel concepts' content positively or negatively. Experiment 1 finds that morpheme-containing labels do have a small but significant effect on the interpretation of novel concepts. Experiment 2 then tests the extent to which these results are explainable merely by the perceived valence of the morphemes in the labels themselves (as opposed to labels influencing the interpretation of novel concepts), producing corroborating evidence that labels change how people conceptualize the novel concept. Experiment 3 modifies Experiment 1 by testing submorphemic labels, in particular velar plosives, and finds no effect on the perceived valence of concepts. Together, these suggest that while labels are something conceptual engineers and other linguistic activists should care about, the effects are subtle and can be overridden by other forms of information.

3. Experiment 1

The goal of Experiment 1 is to test the bluntest tool linguistic activists have at their disposal, namely remixing or coopting known words as a vehicle for novel concepts. This is a common strategy in language design. For example, novel technical uses often *repurpose* existing words in semi-descriptive ways, such as Kuhn's use of "paradigm" to refer to collections of concepts, practices, and methods in science, and the organizational theory use of "wicked" to describe difficult-to-define and open-ended problems. Sometimes, instead, a newly minted label in technical settings combines recognizable roots in novel ways. These include terms for novel technology like "cryptocurrency", "deepfake", or "blockchain" and a quick search in Google Scholar produces examples like "underspecification", "polysystem", and "hyperintensional".

Polysemous coinages and combinations of morphemes allow language designers to piggyback on existing connotations, meanings, and imagery associated with the label to help the audience learn new words (Landes, 2025a; Maitra, 2018). This is not to say that labels need to carry their new meaning literally. Instead, the labels function by evoking certain meanings, values, or modes of presentation. This may occur by framing the described phenomena in a way that embeds them into larger preexisting metaphorical structures (Lakoff, 2004; Thibodeau et al., 2011) or by drawing attention to specific instances or aspects of the defined phenomena. In contrast, if language designers instead use a nonsense label, they have to start building meaning, connotations, imagery, etc., from scratch (Robertson, 1989). Accordingly, Experiment 1 aims to capture the effect labels employing known words have on the perception of the positivity or negativity of novel concepts.

3.1. Materials and methods

To test the effect of label choice on the perception of novel concepts, evocative labels were written for novel concepts. Eight novel concepts were written to share three features (Table 1).[2] First, they were designed to be *unlexicalized*, meaning they do not already correspond to a word in the English lexicon. Second, they were designed to be *neutral* so that they could be framed in either a positive or negative light. This was so that the measured valence was not already at the scale ceiling or floor and different labels had space on the scale to pull the interpretation of the concept either positively or negatively. Third, concepts were designed to be widely and easily *understood*. While impossible to consistently achieve in practice, in theory, they were designed to be the sort of novel concept that causes readers to think, "Oh yeah, that" when seen explained.

For Experiments 1 and 2, one control label was generated for each of the eight concepts. These were designed to be information-less gibberish that nonetheless used letter and sound combinations familiar to English speakers. Control labels were outputs of the pseudoword generator at https://www.randomready.com/fake-words/. Pseudowords with recognizable roots were eliminated, and any recognizable suffixes or prefixes were trimmed from pseudowords. After an initial set of 24 control labels was generated and modified, 15 participants were asked to rate how clear the labels were to pronounce. The highest-scoring eight control labels were then randomly assigned to one of the eight concepts (Table 2).[3]

Each concept was also given four test labels. The test labels were written by hand and designed to highlight or frame certain aspects of the concept. The four test labels were written in a 2×2 design. The first factor was the valence of the framing (positive vs negative features of the concept). The second factor was whether the label was a novel compound noun or a polysemous use of an existing term – that is whether the label gained its framing by combining known words or repurposing single words. The polysemous labels were written to, if possible, use a different part of speech than existing uses of the word. Given the ubiquity of polysemy, this was difficult to achieve, so many of the novel polysemous labels were instead

Table 1. The eight novel concepts written to be unlexicalized, neutral, and easily understood.

Concept	Definition
C1	The action of trying someone else's food at a restaurant to see what it tastes like
C2	The time between when you wake up and when you eat breakfast
C3	Someone who spends a lot of time keeping up with international news
C4	The smell of asphalt or tarmac after rain
C5	A person who never drinks anything caffeinated such as coffee or tea.
C6	When it has been very hot outside during the day and has started cooling off in the evening, and the inside of the building is now warmer than outside
C7	An artist who uses unconventional materials and methods
C8	A person who willingly sleeps less than 5 hours a night

Table 2. The five labels written for each of the eight concepts. The test labels were designed to highlight either positive or negative aspects of the novel concepts, and the control labels were designed to lack any recognizable morpheme.

Concept	Positive-coded compound	Negative-coded compound	Positive-coded polysemy	Negative-coded polysemy	Control
C1	A tastytaste	a greedgrab	a discover	a greed	A ferras
C2	The risingtime	The hungerhour	The rising	The starving	The flockeet
C3	A learnedcitizen	A paranoidreader	A knowledgeable	A worried	A fabran
C4	sweetmist	tarstench	bouquet	stench	bobbon
C5	A strongresolve	a lamehead	A clean	A boring	A struss
C6	sunhug	miseryheat	snuggle	boil	marson
C7	A visionarysupreme	A phonyfraud	A genuine	A pretend	A flunge
C8	A productivehead	A drainedmind	A productive	A drained	A shipoto

chosen because they committed some sort of category mistake when compared to existing senses identified by the Oxford English Dictionary. This resulted in 40 test labels in addition to 8 control labels (Table 2).[4]

The concepts and labels were presented to participants as dictionary-style definitions. For example, in C1 the compound conditions read: "A [*greedgrab/tastytaste*] is the action of trying someone else's food at a restaurant to see what it tastes like". Participants were then asked to rate the valence of the referent: "Is a [greedgrab/tastytaste] a good thing, a bad thing, or somewhere in the middle?" using a 7-point scale (anchored at −3 = bad, 0 = neither good nor bad, 3 = good). The question was written as an object-level question about the referent to better capture participants' understanding of the content of the concept (e.g., what the concept picks out) as opposed to anything meta-linguistically or pragmatically communicated by the definition or label. The scale was centered at 0 to help communicate to participants that the scale was meant to be understood as bipolar (bad to good) as opposed to unipolar (not good to good) (Conrad et al., 2014, pp. 423–424).

Three hypotheses about Experiment 1 were pre-registered:

- (H1): Across all concepts (aggregated), there will be significant differences between the rated valences of different categories.
- (H2): Across all concepts (aggregated), positive-coded compound and positive-coded polysemy will have higher valence scores than control while negative-coded compound and negative-coded polysemy will have lower valence scores than control.
- (H3): For each of the eight concepts, there will be a main effect of category.

Each participant saw all eight concepts in a random order, and the label each participant saw for each concept was randomized. Two hundred and sixty-one participants who spoke English as a first language and were in the US or the UK were recruited via Prolific, and four were excluded for spending less

than 60 s on the survey, leading to a final total of 257. Of the non-excluded participants, 179 (70%) identified as female, 75 (29%) identified as male, and 3 preferred not to say, and the average age was 41.

3.2. Results

Comparing valence scores across label categories (H1), there was a significant difference in a 1-way ANOVA across all concepts ($F(4, 2051) = 24.53$, $p < .001$) (Figure 1). Comparing each of the categories against control (H2), in unpaired unequal variance t-tests with a Bonferroni-corrected alpha of .0125, there was a significant difference between control labels ($M = 0.34$, $SD = 1.38$) and negatively coded polysemy labels ($M = -0.15$, $SD = 1.52$); $t(818.16) = 4.94$, $p < .001$. There was also a significant difference between the control labels and negatively coded compound labels ($M = -0.18$, $SD = 1.52$); $t(807.57) = 5.15$, $p < .001$. Due to the adjusted alpha, there was not a significant difference between control ($M = 0.34$, $SD = 1.38$) and positively coded polysemy ($M = 0.54$, $SD = 1.47$); $t(817.54) = -1.99$, $p = .023$ nor between control and positively coded compound labels ($M = 0.54$, $SD = 1.47$); $t(817.24) = -1.97$, $p = .025$.

Notably, all the results are sitting around the scale midpoint, and the effect size of the ANOVA is small (Cohen's $f = 0.22$). Combining data in this way, however, hides eight underlying patterns, as each concept is unique and has four hand-written labels. Therefore, to test Hypothesis 3, results were broken down by each concept. Analyzing each concept, six of the concepts showed significant differences between categories in 1-way ANOVAs (Figure 1). This includes Concept 1 ($F(4, 252) = 8.94$, $p < .001$), Concept 3 ($F(4, 252) = 8.43$, $p < .001$), Concept 4 ($F(4, 252) = 5.81$, $p = .0017$), Concept 5 ($F(4, 252) = 2.83$, $p = .025$), Concept 6 ($F(4, 252) = 12.16$, $p < .001$), and Concept 7 ($F(4, 252) = 3.01$, $p = .019$). Concept 2 ($F(4, 252) = 0.59$, $p = .067$) and Concept 8 ($F(4, 252) = 2.17$, $p = .073$) showed no significant effect of category. Effect sizes among the significant concepts ranged from small to large (minimum: C2; Cohen's $f = 0.21$, maximum: C6; Cohen's $f = 0.44$).

3.3. Discussion

Novel concepts' labels affect people's initial judgments about the novel concept's content. When paired with a label that frames a novel concept positively or negatively, people's perception of the concept's valence changes. This vindicates the motivation behind linguistic activists' campaigns to introduce revised labels for terms, e.g., "slidkrans" in Swedish, and confirms previous findings that labels are not merely passive linguistic vehicles but influence our perception of their associated content (e.g.,

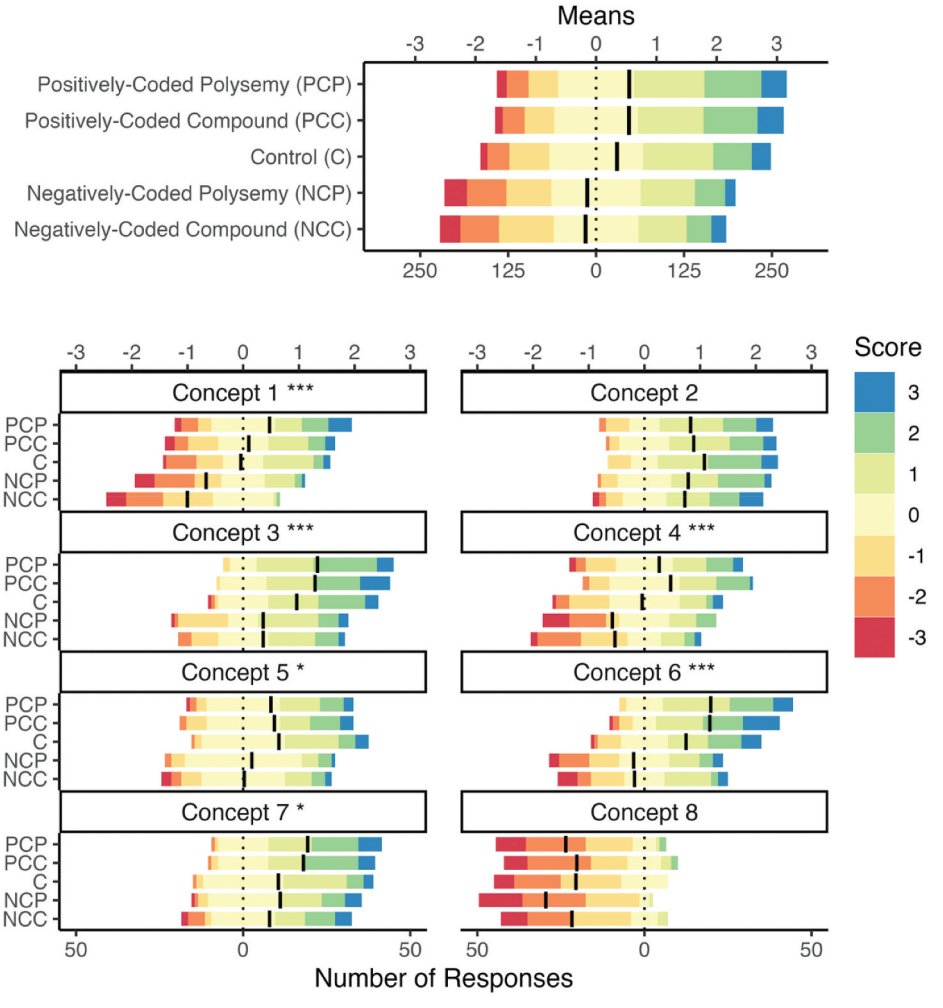


Figure 1. Diverging bar chart showing responses in Experiment 1. The top chart aggregates all concepts broken down by category, and the bottom charts separate results by concept. The width of each bar represents the number of participants who selected the score as a response. All bars are centered around scores of 0. Vertical bars represent the mean score for each condition. The dotted line represents the midpoint of both responses and scale. Asterisks represent the level of significance of ANOVAs ($p \leq .05$ = *; $p \leq .01$ = **, $p \leq .001$ = ***).

Mandelbaum et al., 2024; Pogacar et al., 2021; Wu et al., 2013). Unlike past work, this specifically identified the importance labels play in the projects of activists like conceptual engineers when they are designing novel concepts to spread. It is not just one thing to come up with a new idea – without the right name, the idea risks being interpreted in unintended ways.

Because the stimuli were handwritten and thus idiosyncratic, it is unclear how much to conclude from some of the results. For example, two of the concepts, C2 and C8, did not follow the expected trends. This may be a feature of the concepts themselves – they have the highest and lowest

valence scores in the control condition, respectively. Alternatively, the written labels for C2 and C8 may just fail to adequately frame the concepts in the intended way. While the concepts and labels were workshopped informally, the labels may have nonetheless failed to have the intended framing effect on participants, and this possibility for misjudgment during the design stage speaks to the need for conceptual engineers and other linguistic activists to approach their projects empirically (see Landes, 2025a; Pinder, 2017; Wakil, 2023). Another finding with unclear import is that when all eight concepts are aggregated, the two negative conditions are significantly different from the control, but the two positive conditions fall short of the adjusted alpha of .0125. In unrelated studies, negatively valenced evaluative language has been found to have a larger effect on judgments than positively valenced evaluative language (Baumgartner et al., 2022; Willemsen & Reuter, 2020). The difference observed here may be part of a larger asymmetry between positive and negative information (e.g., Unkelbach et al., 2020), or it may be that the negatively valenced labels just resonated more with participants. The biggest question left unanswered by Experiment 1, however, is what exactly is driving the valence scores?

4. Experiment 2

While Experiment 1 found that labels affect the initial judged valence of novel concepts, Experiment 2 aims to distinguish two distinct but related mechanisms that might be driving the results. The most straightforward interpretation of the findings in Experiment 1 is that the different labels changed how participants interpreted the novel concept's content. For example, calling *the act of trying someone else's food at a restaurant* a "tastytaste" causes people to interpret the communicated act as one that is cute and playful, whereas calling that same act a "greedgrab" causes people to interpret the communicated act as driven by selfishness. An alternative mechanism for the difference in valence scores between labels of a given concept, however, is that the differences are entirely due to the normative denotations and/or connotations of the English words used to construct the labels. Instead of the labels and concepts interacting to produce differences in valence scores, on this alternative, the perceived valences of the novel concepts are directly inherited from the perceived valence of, for example, the words "greed" and "grab".

Many of the terms employed in Experiment 1's labels, such as "greed", "paranoid", and "resolve", express value-laden content in their regular use. Words are typically distinguished as expressing descriptive or evaluative content. Descriptive words like "shelf", "tree", "red", and "large" describe something that does not express any positive or negative judgment. Being red is neither a good nor a bad thing, at least outside of certain specific

contexts. Evaluative terms like "good", "bad", and "wrong" instead express explicitly value-laden content. The aforementioned terms from Study 1 like "greed", "paranoid", and "resolve" are *thick evaluative terms* expressing *thick concepts*, or concepts that contain both descriptive and evaluative content (Väyrynen, 2025; Williams, 1985). Thick concepts like *frugal*, *thrifty*, and *economical* require more than fulfilling the descriptive features of being careful about spending money. *Frugal*, *thrifty*, and *economical* also carry with them an evaluation that they are careful about spending *in a good way* – otherwise more apt concepts would be *cheap*, *stingy*, or *tightfisted*. Importantly, the descriptive and evaluative aspects of thick terms and concepts are inexorably connected. The evaluative aspect of thick terms like "stingy" cannot be canceled out conversationally, and it is contradictory to emphasize that someone is stingy but in a good way (Baumgartner et al., 2024; Willemsen & Reuter, 2021).

Because the descriptive and evaluative content of thick terms and thick concepts are necessarily connected, the evaluative content contained in the constituents of Experiment 1's labels (including the polysemous labels, see Fischer, (2020)) almost certainly had an effect on valence scores. Indeed, much of the intended framing in Experiment 1 was achieved using thick concepts, including C1's "a *greed*grab", C2's "the *starving*", C3's "a *paranoid*reader", and C4's "*stench*". The question is whether constitutive words' standing evaluative content was the only driver of positive or negative valence scores. Perhaps, the observed negative valence for "greedgrab" was just the approximate combined negative valence of the thick evaluative concept "greed" (and plausibly "grab"), and not evidence that the label affected the interpretation of the definition.

Some of the labels in Experiment 1 do not contain evaluative concepts, e.g., "discover" (C1) and "sunhug" (C6), but this is not necessarily a problem for the theory that labels are the sole driver of valence scores. While "hug" does not carry evaluative content – a "bad hug" is not an oxymoron in the way a "good stench" is – "hug" is still a *value-associated term* (Reuter et al., 2022). For many people, hugs are usually a nice thing, so the idea of the sun hugging us that is evoked by "sunhug" nonetheless carries positive associations even if that positivity is only contingent. In this way, the results of Experiment 1 may merely reflect the inherent and associated valence of the coopted English words, rather than the result of labels influencing how people interpret and understand the novel concepts.

4.1. Materials and methods[5]

Experiment 2 tests the possibility that the scores in Experiment 1 were driven by judgments about the valence of the words used to label the new concepts as opposed to the labels influencing understanding of the content

of the novel concept. To test this, Experiment 2 compares the valence scores of the labels alone vs the labels alongside the definition. If the evaluative language used in the labels is the primary driver of valence scores in Experiment 1, there should be no difference in valence scores with and without the definition. If, however, the labels and definitions are interacting, because the concepts were written to be neutral and the labels were written to be valenced, the addition of the definition should moderate valence scores and bring them toward the midpoint – specifically the valence observed with the control label in Experiment 1.

This only works for the labels that drove valence scores toward the ends of the scale in the intended direction in Experiment 1. Accordingly, not all the label-concept pairs from Experiment 1 are suitable for Experiment 2, as some label-concept pairs had non-significant or unexpected effects on valence. Therefore, two exclusion criteria were applied to the concepts. First, concepts without significant ANOVAs were excluded, eliminating C2 and C8. Second, concepts that did not have at least one positive-coded term above the control label and one negative-coded term below the control label were excluded, eliminating C5. From the remaining five concepts, one positive label with a mean valence score above that concept's control label and one negative label with a mean valence score below that concept's control label were chosen (Table 3).

The stimuli of Experiment 1 were then modified to allow for the definition to be added or removed. The first part of the stimuli, seen by all participants, was a short sentence containing the label. These were meant to contain very little information about the meaning of the novel concept while being correct usages of the novel concept. To feel natural, however, the sentences unavoidably included some inferable information about the meaning of the novel concept (see Landes, 2025a, Section 3). The second part of the stimuli, only seen by half of the participants, was an explanation of the meaning of the novel concept using the definitions from Experiment 1.

Participants were assigned to one of the four groups in a 2×2 between-subject design: positive label with definition, positive label only, negative label with definition, or negative label only. Each participant saw all five concepts for that condition in a counterbalanced order. Two hundred and forty-one participants who spoke English as their first language and were in the US or the UK were recruited through Prolific. Of these participants, 147 (61%) identified as female, 93 (38%) identified as male, and 3 preferred not to say. The average age of participants was 39.

Two hypotheses were pre-registered about Experiment 2:

Table 3. The concepts used for Experiment 2. Each participant saw either a positively coded or negatively coded label introduction. Only participants in the definition condition saw the definition, which was the same for both positively coded and negatively coded labels.

Concept	Label introduction	Definition
C1	At lunch, I had a [discover/greedgrab]	– by that I mean I tried someone else's food at a restaurant to see what it tastes like
C3	John is a [knowledgeable/worried]	– by that I mean he is someone who spends a lot of time keeping up with international news
C4	I woke up this morning to [sweetmist/tarstench]	– by that I mean the smell of asphalt or tarmac after rain
C6	Right now my apartment is [snuggle/boil]	– by that I mean the hot weather during the day warmed the inside up, but now that the weather has cooled off in the evening, the inside of the building is now warmer than the outside
C7	Jose is a [genuine/phonyfraud]	– by that I mean he is an artist who uses unconventional methods and materials in his artwork

- (H4): Valence scores of positive-coded labels without clarifying definitions will not equal the valence scores of the labels with clarifying definitions.
- (H5): Valence scores of negative-coded labels without clarifying definitions will not equal the valence scores of the labels with clarifying definitions.

4.2. Results

In a two-tailed unpaired unequal variance t-test,[6] there was a significant difference between positively-coded labels with definitions (M = 1.06, SD = 1.42) and positively coded labels without definitions (M = 1.34, SD = 1.40); t(602.5) = 2.42, p = .016. The effect size was small (Cohen's d = 0.20). Similarly, there was a significant difference between negatively-coded labels with definitions (M = −.09, SD = 1.77) and negatively coded labels without definitions (M = −1.2, SD = 1.36); t(559.58) = 8.61, p < .001). The effect size was medium (Cohen's d = 0.70). An exploratory analysis breaking down results by concept-label pair similarly found that the definition scores were significantly closer to the midpoint in seven of the 10 pairs, although one of the sets (C1 Positive) went in the opposite direction as predicted by H4 (Figure 2 and Table 4).

4.3. Discussion

The inclusion of the definition moved mean valence scores toward the midpoint, meaning the perceived valence is stronger when the label is by itself. This occurred when all five concepts were aggregated and, when broken down, in seven of the 10 conditions. In conjunction with Experiment 1, the findings demonstrate that both labels and concepts affect, but do not completely determine, the perceived valence of the concept. The observed effect of valence scores in Experiment 1 is more than participants' judgments of the valence of thick terms and value-associated terms used in the labels.

Participants are therefore using both the labels and the definitions as complementary sources of information to determine a novel concept's content. This is in line with work in marketing that found that the gender-coding of names affects the perceived warmth of companies, and consequently, consumer choices (Pogacar et al., 2021). If anything, this suggests the connection between marketing and linguistic activism runs deep. In the same way that a brand name is the first point of contact of a consumer with a company (Robertson, 1989), a label is a person's first point of contact with a designed concept. It will color how they learn the concept, even if explicit definitions are provided. The question now is whether this is limited to

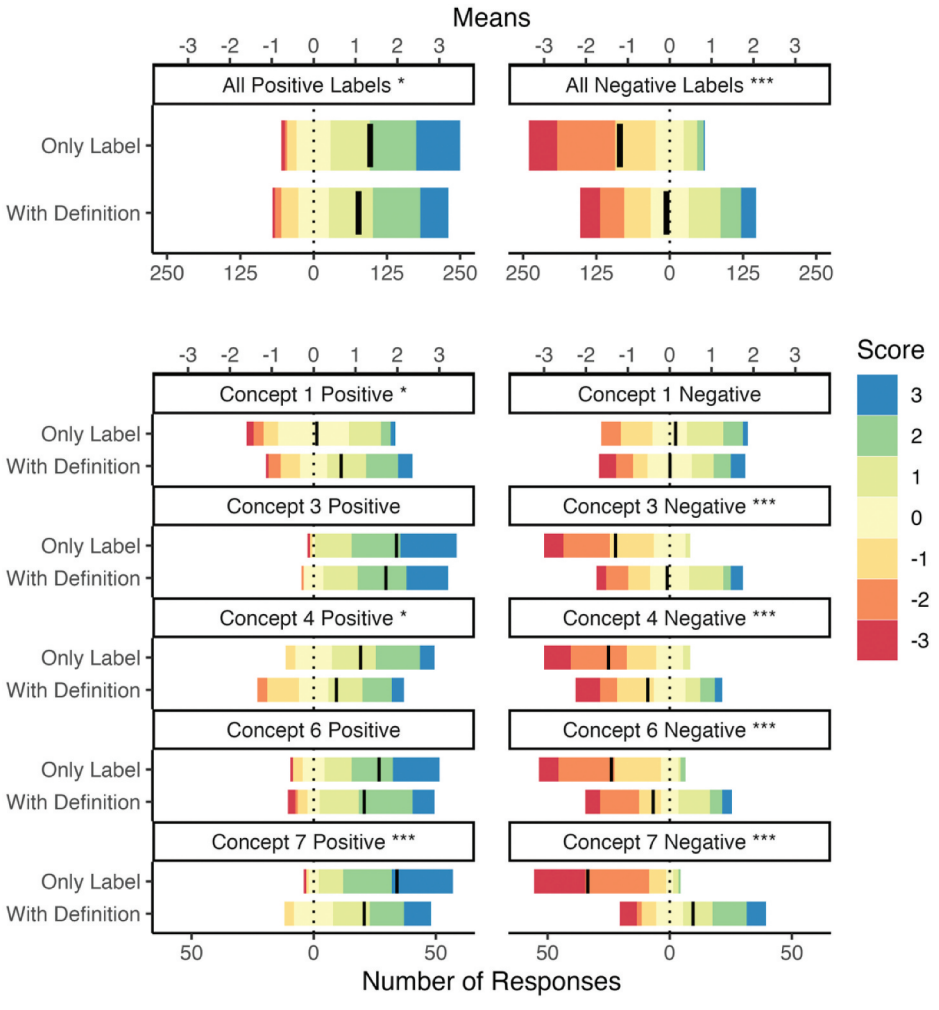


Figure 2. Diverging bar chart showing results of Experiment 2. The top chart shows all responses while the bottom chart breaks responses down by concepts. The width of each bar represents the number of participants who selected the score as a response. All bars are centered around scores of 0. Vertical bars represent the mean score for each condition. The dotted line represents the midpoint of both responses and scale. Asterisks represent the level of significance of t-tests ($p \leq .05$ = *; $p \leq .01$ = **, $p \leq .001$ = ***).

obvious information or whether more subtle submorphemic cues also color the interpretation of novel concepts.

5. Experiment 3

Experiment 1 and Experiment 2 tested the role of morphemic information, that is, information at the level of discrete meanings. C4's tarstench has two morphemes – tar and stench – each of which carries apparent semantic information. Moreover, while "tarstench" is not in the English lexicon, the

Table 4. Results of two-tailed unpaired unequal variance t-tests in Experiment 2, broken down by condition and concept.

Concept	Label Framing	Definition Condition	Mean	sd	t	df	P value
C1	Positive	With Definition	0.65	1.51	2.31	114.86	
		Only Label	.066	1.26			
	Negative	With Definition	0	1.79	0.46	110.19	=.65
		Only Label	0.13	1.36			
C3	Positive	With Definition	1.72	1.12	1.23	118.98	=.22
		Only Label	1.97	1.12			
	Negative	With Definition	- .067	1.60	4.97	102.43	
		Only Label	−1.30	1.06			
C4	Positive	With Definition	0.53	1.42	2.52	111	
		Only Label	1.11	1.10			
	Negative	With Definition	−0.53	1.71	3.51	102.82	
		Only Label	−1.47	1.42			
C6	Positive	With Definition	1.20	1.50	1.36	117.83	=0.18
		Only Label	1.56	1.38			
	Negative	With Definition	− 0.4	1.71	3.68	99.62	
		Only Label	− 1.4	1.12			
C7	Positive	With Definition	1.20	1.22	3.58	118.81	
		Only Label	1.98	1.19			
	Negative	With Definition	0.55	1.84	9.07	97.54	
		Only Label	− 1.97	1.12			

nearby noun phrase "tar stench" is a valid construction, albeit one that is likely not naturally interpreted as meaning the *smell of asphalt or tarmac after rain*. Experiment 3 instead aims to explore whether subtler sub-morphemic information also affects the interpretation of concepts,[7] building upon Mandelbaum et al. (2024) study of velar plosives (k, g).

Mandelbaum et al. found that participants were more likely to pair pseudowords with velar plosives than their counterparts with supplied negative definitions than supplied positive definitions. They also found that when participants were asked whether they thought the pseudowords had a negative or positive definition (in the abstract), participants were more likely to guess the pseudowords had a negative meaning. This suggests that sound alone is enough information for participants to draw information about the meaning of words (see also Gatti et al., 2024), and phonesthemes are important to consider during language design because they affect people's inferences when they have no other information to draw from. What remains unanswered, however, is whether phonesthemes affect perceptions about content when the content is explicitly supplied. Therefore, Experiment 3 will use the pseudowords from Mandelbaum et al. (2024) to test whether phonesthemes influence understandings of novel conceptual content. As opposed to Mandelbaum et al., who tested how velar plosives affected guesses about pseudowords' definition, Experiment 3 tests whether velar plosives affect the interpretation of *supplied* definitions in the way seen in Experiment 1.

5.1. Materials and methods[8]

Using all eight concepts from Experiment 1, Experiment 3 assigns each concept a pair of pseudowords from Mandelbaum et al. (2024). The pairs are identical except that one contains a velar plosive and the other contains another sound. Eight of Mandelbaum et al.'s pseudoword pairs were randomly assigned to each of the eight concepts used in Experiment 1 (Table 5). Otherwise, the stimuli and test questions were identical to Experiment 1. Participants were shown a dictionary-style definition of the new concept and asked to rate the valence of the item, with the label assigned to the concept changing based on condition.

The design had two conditions (Velar Plosive/Control). Each participant saw all eight concepts, but the condition (Velar Plosive/Control) was randomized for each participant for each concept. Through Prolific, 150 participants who spoke English as a first language and were in the US or the UK were recruited. The average age of participants was 39. Ninety-three participants (62%) identified as female, 54 (36%) identified as male, and 5 preferred not to say.

One hypothesis was pre-registered:

- (H6): Concepts with labels with velar plosives will be rated more negatively than the same concepts with labels without velar plosives.

5.2. Results

Combining all eight concepts, there was no difference between types of labels (Figure 3). In a two-tailed unpaired t-test, the Velar Plosive labels ($M = 0.26$, $SD = 1.49$) had nearly identical means to the control labels ($M = 0.25$, $SD = 1.49$); $t(1198) = 14$, $p = 0.89$. An exploratory analysis of each concept similarly found no significantly different pairs, although C8 was trending ($t(148) = 1.77$, $p = .079$). Therefore, there is no evidence for H6.

Table 5. The label pairs for each of the eight concepts in Experiment 3. The labels are a randomized subset from Mandelbaum et al. (2024).

Concept	Velar Plosives	Control
C1	vlack	vlarp
C2	yeckling	yeppling
C3	huckey	hurney
C4	yeckuse	yelluse
C5	reckol	renfol
C6	fepick	fepish
C7	konil	honil
C8	holck	holst

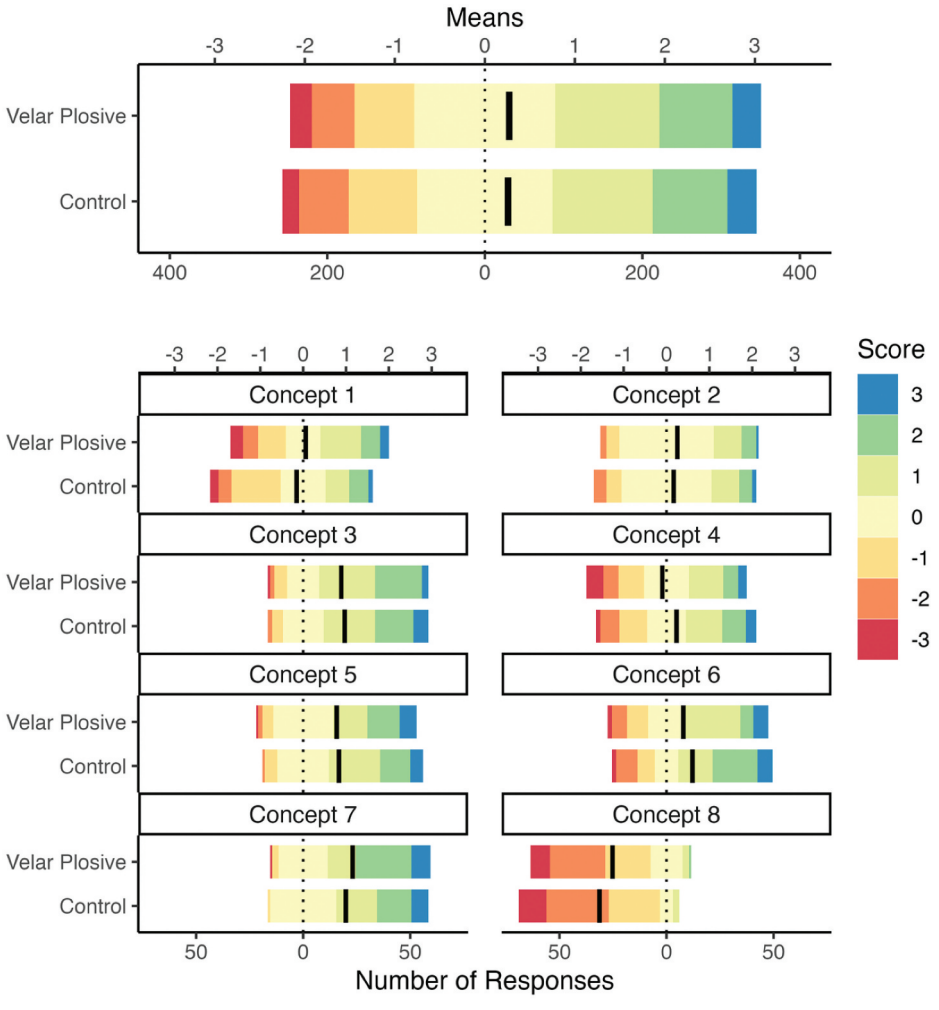


Figure 3. Diverging bar chart showing results of Experiment 3. The top chart shows all responses, while the bottom charts break responses down by concepts. The width of each bar represents the number of participants who selected the score as a response. All bars are centered around scores of 0. Vertical bars represent the mean score for each condition. The dotted line represents the midpoint of both responses and the scale. Asterisks represent the level of significance of t-tests ($p \leq .05$ = *; $p \leq .01$ = **, $p \leq .001$ = ***).

5.3. Discussion

Including velar plosives in labels does not affect the perceived valence of novel concepts in the same way as morphemes. Experiment 3 used labels previously found to license inferences that a word has a negative definition (Mandelbaum et al., 2024) and used them in a design similar to Experiment 1. While perceived valence was affected – albeit inconsistently – in Experiments 1 and 2 by labels containing known English words,

Experiment 3 failed to find a connection between submorphemic information in labels and the valence of corresponding content.

The difference between Experiment 3 and Mandelbaum et al. licenses a tentative conclusion about how submorphemic information like velar plosives affects perceptions of content. In one task, Mandelbaum et al. asked participants to guess definitions from a supplied set, and in another task, they had participants guess, in the abstract, whether the word had a positive or negative definition see also Perlman et al. (2015), Sapir (1929), Wu et al. (2013). Experiment 3 instead provided participants with explicit definitions for the word. The difference in findings suggests that while whole-word morphemic information has a strong enough effect on interpretation to influence how people understand explicit definitions, submorphemic information is not strong enough to do so, at least initially.

Therefore, language designers may only need to worry about submorphemic information when they do not expect people to encounter explicit explanations of new concepts. Campaigns designed around high-profile, explicit definitions may be safe from submorphemic information, while campaigns relying on organic spread between language users need to take more care to avoid forms of sound symbolism and phonesthemes. This is because when people learn a new word from hearing another person using it, they need to infer the meaning of the word from things like context and sound symbolism, as opposed to learning the meaning from definitions.

Notice, however, that it is unclear how generalizable Experiment 3's results are beyond English. While many patterns of sound symbolism are nearly universal (Ćwiek et al., 2021; Shinohara & Kawahara, 2010), there are nonetheless still subtle differences and exceptions in sound symbolic meaning across languages (Shih et al., 2019; Thompson et al., 2021). More importantly, sound symbolism plays a much smaller role in English than in many other languages, such as Japanese (Imai et al., 2008; Kita, 1997). Language designers in other languages would therefore be prudent to reflect on whether their language may behave differently.

6. General discussion

Language designers should pay attention to the labels they use to express novel meanings or concepts. Language design, such as conceptual engineering, involves the intentional change of language to address some hermeneutical gap or other inadequacy and requires teaching people a new or revised concept to bring about that change. Despite what language designers might wish, they do not have direct control of people's linguistic beliefs or use patterns, meaning whatever language designers attempt to spread is liable to be interpreted in unintended ways. Understanding the factors that drive perceptions of meaning or

conceptual content – both synchronically and diachronically – will help language designers fulfill their goals.

The above experiments looked at the effect of labels on synchronic conceptual uptake. Experiment 1 examined whether recognizable roots, both in novel compound nouns and in novel polysemous uses, affect the valence of the perceived content, finding that it does. Experiment 2 examined whether these results were driven solely by the evaluative content (or associated values) of the labels themselves and found that the valence scores of labels alone were more extreme than the valence scores of labels with definitions. Therefore, labels change how people *interpret* definitions and are not passive parts of linguistic activism (Koch, 2024; Koslow, 2022). When this active role is well-harnessed, careful remixing of existing language can help serve as an additional epistemic tool during linguistic activism that helps people pick up the desired changes in linguistic beliefs and conceptual content (Landes, 2025a, 2025b). By giving labels names that are evocative of the desired associated meaning or concept, people learning the words can use those labels to speed up the process of learning the new idea.[9]

Whether language designers aiming to deploy labels in responsible and ethical ways choose to pair their concept with an evocative or meaningless label should accordingly depend on their goals and real-world conditions. Language designers who care about the speed of spread and who want to responsibly exploit the sort of epistemic processes observed here may choose to purposefully use valenced, morpheme-containing labels. This will help target language users infer the intended content of the novel concept when they hear it used by others in everyday conversation at the cost of raising the risk of misinterpretation and semantic drift. Language designers who are instead more concerned with the deliberate and careful communication of specific content may instead opt to use nonsense labels. When accuracy is paramount, it might be best to hedge one's bets and call *the act of trying someone else's food at a restaurant* a "ferras" instead of risking unexpected influences from labels like "tastytaste" or "greedgrab".

Finally, how do we make sense of the difference in morpheme vs submorphemic information found in Experiments 1 and 3? During language learning, labels are one source of information among many. We can also use context clues from utterances containing the novel lexical item as well as explicit sources of information such as dictionaries or campaigns that spread the lexical item (MacGregor et al., 2021; Milles, 2011b). In low-context settings, people encountering novel words may draw upon subtle sources of information like phonesthemes because it is the only information they have at their disposal. However, as more clear-cut sources of information become available – such as explicit definitions – label-level information may be relied upon less. This is rational since while there are regularities

between label and meaning, these regularities are not perfect. Words like "good", "golden", "lucky", and "kind" contain velar plosives but have positive meanings, and misnomers like "strawberry" (which is not actually a berry) or the icy "Greenland" mean even apparent compound nouns cannot be trusted.

Additional information does not wash out all types of information contained in the label equally, explaining the differences between Experiments 1 and 3. More research is needed, however, to identify why. Differences may be driven by metacognitive factors (morphemic information is likely easier to process than submorphemic information), epistemic factors (in English, morphemic information is likely a more reliable guide to associated content than sound-symbolic cues), or something else. It is also unclear what cognitive mechanism leads to submorphemic information's inefficacy. One possibility is that submorphemic information is largely ignored, except when needed. Another possibility is that submorphemic information registers with language users but is suppressed or overridden by more explicit and reliable explanations of content. A combination of eye-tracking and reaction time studies may be able to uncover the mechanisms at play (for an introduction, see Fischer et al. 2024). Eye-tracking paradigms (e.g., Hyönä et al., 2020) would reveal which sources of information participants focus on when encountering novel word-concept pairs in context. Reaction-time studies could reveal if submorphemic information is still processed when explicit definitions are available by testing if increased time is spent on a task when submorphemic information is incongruous with the definition – for example, when a positive concept is attached to a label containing velar plosives (see Lupyan & Casasanto, 2015).

What cannot be ruled out, however, is that submorphemic information has more subtle effects in high-information environments. For one, subtle cues may, over time, affect things like perceived valence. Mandelbaum et al. (2024) found that the presence of velar plosives predicted the perceived valence of *established* English words, so phonesthemes and related phenomena may instead slowly influence valence over the course of decades or generations. Indeed, less subtle forms of information contained in labels seem to have a continuing effect on perceptions of referents. For example, in the UK one occasionally hears the complaint that the Labour Party has forgotten what its name is (i.e., that it is the party of labor). Even though UK voters have tons of information about the Labour Party's policies and members, UK voters still nonetheless expect conformity between the name of the party and its policies. Indeed, *all* labels – submorphemic and morphemic – may continuously pressure the associated content to conform to the labels, driving enduring conceptual change through processes like semantic drift. Further experiments, however, are needed to understand how labels' influence plays out over time.

7. Conclusion

Novel concepts need labels. Sometimes these novel concepts, like *global warming*, *social distancing*, and *the Anthropocene*, are important for coordinating action. Sometimes these novel concepts, like *neurodivergence*, *snippa*, and *mansplaining*, are important for addressing structural or hermeneutic injustices. Other times we just need a name for a novel varietal of tomatoes or our newest dusty academic theory. This paper aimed to establish the importance of taking care when deciding what label will serve as the novel concept's vehicle, focusing on the effect labels have on people's initial perceptions of concepts' valence. Not only does the label assigned to a novel concept affect people's perceptions of whether the goodness or badness of the thing captured by the concept, but this effect appears to be due to the way the label and novel concept interact in people's minds, as the effect of the label is strongest in low context settings without the supplied definition. Not all information contained in the label is relevant all the time, however. In cases where the information contained in the label is submorphemic and subtle, this paper failed to find the effect of labels on the perceived valence of novel concepts. Accordingly, language designers such as conceptual engineers need to pay careful attention to what they call their idea, otherwise their creation risks being interpreted in unintended ways, threatening to undermine the motivation for introducing it in the first place.

Notes

1. I would like to thank Fenner Tanswell, Kalle Rocklinger, Kevin Reuter, Manuel Gustavo Isaac, Kathryn Francis, Eugen Fischer, Lucien Baumgartner, James Andow, and anonymous referees for helpful feedback and discussions about this project. I would especially like to thank Pascale Willemsen for suggesting that label influence could be operationalized with valence scores. This work was supported by the Swiss National Science Foundation under grant number 181,082.
2. While this paper uses the language of "novel concepts," this choice was made to better situate the study within the conceptual engineering literature. I see no reason why the results cannot be interpreted as studying labels' effect on the perception of *novel definitions* or *novel words*.
3. Due to the experimental design, it cannot be ruled out that information such as phonographic systematicity or iconicity in the control pseudoword labels influenced participants' interpretation of the concepts in a way that affects results. However, given Experiment 3's null results using pseudowords with known submorphemic information, this possibility is unlikely.
4. All materials and data for this paper can be found at the online repository: https://osf.io/kw7ua
5. All materials and data for this paper can be found at the online repository: https://osf.io/kw7ua
6. An initial check of the data using f tests found the comparison between negative valence with/without definitions violated the assumption of equal variance ($f = 1.71$,

num df = 299, denom df = 299, $p < .001$). Therefore, unequal variance t-tests were used throughout analysis of Experiment 2. It had no effect on significance in any reported test.

7. Whether phonesthemes are submorphemes or morphemes is a point of contention (Svantesson, 2017, pp. 6–7). The larger motivation behind this study stands regardless of how one comes down on the distinction because English phonesthemes are far subtler cues than those used in Experiments 1 and 2.
8. All materials and data for this paper can be found at the online repository: https://osf.io/kw7ua
9. Whether this evocation is part of concepts' content or merely a tool for communicating concepts' content will depend on how drastically the effect of the evocation differs from the content communicated about the idea (as well as one's philosophical commitments about the nature of content, including the relationship between content and anchoring events such as baptisms).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung [181082].

ORCID

Ethan Landes http://orcid.org/0000-0002-1186-1717

References

Argo, J. J., Popa, M., & Smith, M. C. (2010). The sound of brands. *Journal of Marketing*, *74* (4), 97–109. https://doi.org/10.1509/jmkg.74.4.097

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex, A Journal Devoted to the Study of the Nervous System and Behavior*, *63*, 196–205. https://doi.org/10.1016/j.cortex.2014.08.025

Aveyard, M. E. (2012). Some consonants sound curvy: Effects of sound symbolism on object recognition. *Memory and Cognition*, *40*(1), 83–92. https://doi.org/10.3758/s13421-011-0139-3

Baumgartner, L., Willemsen, P., & Reuter, K. (2022). The polarity effect of evaluative language. *Philosophical Psychology*, *37*(8), 1–18. https://doi.org/10.1080/09515089.2022.2123311

Baumgartner, L., Willemsen, P., & Reuter, K. (2024). The polarity effect of evaluative language. *Philosophical Psychology*, *37*(8), 2158–2175. https://doi.org/10.1080/09515089.2022.2123311

Baus, C., Carreiras, M., & Emmorey, K. (2013). When does iconicity in sign language matter? *Language and Cognitive Processes*, *28*(3), 261–271. https://doi.org/10.1080/01690965.2011.620374

Cappelen, H. (2018). *Fixing language*. Oxford University Press. https://doi.org/10.1093/oso/9780198814719.001.0001

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, *30*(3), 348–369. https://doi.org/10.1016/0749-596X(91)90041-H

Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin and Review*, *8*(3), 519–523. https://doi.org/10.3758/BF03196187

Chalmers, D. J. (2020). What is conceptual engineering and what should it be? *Inquiry*, 1–18. https://doi.org/10.1080/0020174X.2020.1817141

Conrad, F. G., Schober, M. F., & Schwarz, N. (2014). Pragmatic processes in survey interviewing. In T. M. Holtgraves (Ed.), *The Oxford Handbook of language and social psychology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199838639.013.005

Cooper, R. L. (1990). *Language planning and Social change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511620812

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Ünal-Logacev, Ö., & Winter, B. (2021). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*, (1841), 20200390. https://doi.org/10.1098/rstb.2020.0390

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. https://doi.org/10.1016/j.tics.2015.07.013

Dingemanse, M., & Thompson, B. (2020). Playful iconicity: Structural markedness underlies the relation between funniness and iconicity. *Language and Cognition*, *12*(1), 203–224. https://doi.org/10.1017/langcog.2019.49

Fischer, E. (2020). Conceptual control: On the feasibility of conceptual engineering. *Inquiry*, 1–29. https://doi.org/10.1080/0020174X.2020.1773309

Fischer, E., Engelhardt, P. E., Bauer, A. M., Alfano, M., Allard, A., Baumgartner, L., Cova, F., Engelhardt, P., Fischer, E., Meyer, H., Reuter, K., & Sytsma, J. (2024). Psycholinguistic Experiments: A case study on default inferences in philosophical arguments-analysing the argument from illusion. In S. Kornmesser, K. Thompson, & M. Wyszynski (Eds.), *Experimental philosophy for beginners: A gentle introduction to methods and tools* (pp. 355–436). Springer International Publishing. https://doi.org/10.1007/978-3-031-58049-9_8

Fricker, M. (2007). *Epistemic injustice: power and the ethics of knowing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237907.003.0008.

Gatti, D., Rodio, F., Rinaldi, L., & Marelli, M. (2024). On humans' (explicit) intuitions about the meaning of novel words. *Cognition*, *251*, 105882. https://doi.org/10.1016/j.cognition.2024.105882

Grote, K., & Linz, E. (2003). The influence of sign language iconicity on semantic conceptualization. In W. G. Müller & O. Fischer (Eds.), *From sign to signing* (Vol. 3, *Iconicity in Language and Literature*, pp. 23–40). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/ill.3.05gro

Haslanger, S. (2000). Gender and race: (what) are they? (what) do we want them to Be? *Noûs*, *34*(1), 31–55. https://doi.org/10.1111/0029-4624.00201

Hyönä, J., Pollatsek, A., Koski, M., & Olkoniemi, H. (2020). An eye-tracking study of reading long and short novel and lexicalized compound words. *Journal of Eye Movement Research*, *13*(4). https://doi.org/10.16910/jemr.13.4.3

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Isaac, M. G., Koch, S., & Nefdt, R. (2022). Conceptual engineering: A road map to practice. *Philosophy Compass*, *17*(10), e12879. https://doi.org/10.1111/phc3.12879

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Ling*, *35*(2), 379–416. https://doi.org/10.1515/ling.1997.35.2.379

Klink, R. R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters*, *11*(1), 5–20. https://doi.org/10.1023/A:1008184423824

Koch, S. (2024). How words matter: A psycholinguistic argument for meaning revision. *Mind and Language*, *39*(3), 364–380. https://doi.org/10.1111/mila.12494

Koslow, A. (2022). Meaning change and changing meaning. *Synthese*, *200*(2), 94. https://doi.org/10.1007/s11229-022-03563-8

Lakoff, G. (2004). *Don't think of an elephant! know your values and frame the debate: The essential guide for progressives*. Chelsea Green Pub. Co.

Landes, E. (2025a). Conceptual engineering should be empirical. *Erkenntnis*. https://doi.org/10.1007/s10670-025-00923-x

Landes, E. (2025b). How language teaches and misleads: 'coronavirus' and 'Social distancing' as case studies. In M. G. Isaac, S. Koch, & K. Scharp (Eds.), *New perspectives on conceptual engineering*. Synthese Library.

Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition*, *7*(2), 167–193. https://doi.org/10.1017/langcog.2014.21

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083. https://doi.org/10.1111/j.1467-9280.2007.02028.x

MacGregor, C., Petersen, A., & Parker, C. (2021). Promoting a healthier, younger you: The media marketing of anti-ageing superfoods. *Journal of Consumer Culture*, *21*(2), 164–179. https://doi.org/10.1177/1469540518773825

Maitra, I. (2018). New words for old wrongs. *Episteme*, *15*(3), 345–362. https://doi.org/10.1017/epi.2018.27

Majnemer, J., & Meibauer, G. (2023). Names from nowhere? Fictitious country names in survey vignettes affect Experimental results. *International Studies Quarterly*, *67*(1), sqac081. https://doi.org/10.1093/isq/sqac081

Mandelbaum, E., Ware, J., & Young, S. (2024). The sound of slurs: Bad sounds for bad words. In S. Nichols & J. Knobe (Eds.), *Oxford studies in experimental philosophy* (Vol. 5, pp. 221–246). Oxford University Press. https://doi.org/10.1093/9780198918905.003.0010

Milles, K. (2011a). Feminist language planning in Sweden. *Current Issues in Language Planning*, *12*(1), 21–33. https://doi.org/10.1080/14664208.2011.541388

Milles, K. (2011b). 'Snippa'. *Gender and Language*. *5*(1), 89–109. https://doi.org/10.1558/genl.v5i1.89

Milles, K., Holmdahl, J., Melander, I., & Fugl-Meyer, K. (2018). 'Something that stretches during sex': Replacing the word hymen with vaginal corona to challenge patriarchal views on virginity. *Gender and Language*, *12*(3), 294–317. https://doi.org/10.1558/genl.31894

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*(4), 259–305. https://doi.org/10.1016/j.cogpsych.2006.12.001

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, (1651), 1651), 20130299. https://doi.org/10.1098/rstb.2013.0299

Neufeld, E. (in press). Engineering Social concepts: Labels and the Science of categorization. In S. Haslanger, K. Jones, G. Restall, F. Schroeter, & L. Schroeter (Eds.), *Mind, language, and Social Hierarchy: Constructing a shared Social world*. Oxford University Press. https://philarchive.org/rec/NEUESC-2

Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, *28*(1), 225–252. https://doi.org/10.1146/annurev.anthro.28.1.225

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*(2), 173–186. https://doi.org/10.1016/j.jecp.2012.05.004

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, *22*(11), 1419–1421. https://doi.org/10.1177/0956797611421791

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, *2*(8), 150152. https://doi.org/10.1098/rsos.150152

Pinder, M. (2017). Does Experimental philosophy have a role to play in Carnapian explication? *Ratio*, *30*(4), 443–461. https://doi.org/10.1111/rati.12164

Pogacar, R., Angle, J., Lowrey, T. M., Shrum, L. J., & Kardes, F. R. (2021). Is nestlé a lady? The feminine brand name advantage. *Journal of Marketing*, *85*(6), 101–117. https://doi.org/10.1177/0022242921993060

Reuter, K., Baumgartner, L., & Willemsen, P. (2022). Tracing thick and thin concepts through corpora. *Language and Cognition*, *16*(2), 263–282. https://doi.org/10.1017/langcog.2023.35

RFSU. (2017, December 12). *Vaginal Corona*. https://www.rfsu.se/om-rfsu/om-oss/in-english/national-work/sex-and-relations/the-vaginal-corona/

Robertson, K. (1989). Strategically desirable brand name characteristics. *Journal of Consumer Marketing*, *6*(4), 61–71. https://doi.org/10.1108/EUM0000000002563

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, *12*(3), 225–239. https://doi.org/10.1037/h0070931

Saussure de, F. (W. Baskin, P. Meisel, & H. Saussy (Eds.). (1916). *Course in general linguistics*. Columbia University Press.

Shih, S. S., Ackerman, J., Hermalin, N., Inkelas, S., Jang, H., Johnson, J., & Yu, A. (2019). Cross-linguistic and language-specific sound symbolism: Pokémonastics. In *Unpublished manuscript*. University of Southern California, University of California, Berkeley, Keio University, National University of Singapore and University of Chicago.

Shinohara, K., & Kawahara, S. (2010). A cross-linguistic study of sound symbolism: The images of size. *Annual Meeting of the Berkeley Linguistics Society*, *36*(1), 396–410. https://doi.org/10.3765/bls.v36i1.3926

Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin and Review*, *25*(5), 1619–1643. https://doi.org/10.3758/s13423-017-1361-1

Svantesson, J.-O. (2017). Sound symbolism: The role of word sound in meaning. *WIREs Cognitive Science*, *8*(5), e1441. https://doi.org/10.1002/wcs.1441

Thibodeau, P. H., Boroditsky, L., & Lauwereyns, J. (2011). Metaphors we think with: The role of metaphor in reasoning. *Public Library of Science ONE*, *6*(2), e16782. https://doi.org/10.1371/journal.pone.0016782

Thompson, A. L., Hoey, T. V., & Do, Y. (2021). Articulatory features of phonemes pattern to iconic meanings: Evidence from cross-linguistic ideophones. *Cognitive Linguistics*, *32*(4), 563–608. https://doi.org/10.1515/cog-2020-0055

Unkelbach, C., Alves, H., & Koch, A. (2020). Negativity bias, positivity bias, and valence asymmetries: Explaining the differential processing of positive and negative information. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 62, pp. 115–187). Academic Press. https://doi.org/10.1016/bs.aesp.2020.04.005

Väyrynen, P. (2025). Thick ethical concepts. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2025). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/thick-ethical-concepts/

Wakil, S. (2023). Experimental explications for conceptual engineering. *Erkenntnis*, *88*(4), 1509–1531. https://doi.org/10.1007/s10670-021-00413-w

Westbury, C., & Hollis, G. (2019). Wriggly, squiffy, lummox, and boobs: What makes some words funny? *Journal of Experimental Psychology: General*, *148*(1), 97–123. https://doi.org/10.1037/xge0000467

Willemsen, P., & Reuter, K. (2020). Separability and the effect of valence: An empirical study of thick concepts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *42*. https://escholarship.org/uc/item/3gc9c236.

Willemsen, P., & Reuter, K. (2021). Separating the evaluative from the descriptive: An empirical study of thick concepts. *Thought: A Journal of Philosophy*, *10*(2), 135–146. https://doi.org/10.1002/tht3.488

Williams, B. (1985). *Ethics and the limits of philosophy*. Fontana Press.

Wu, L., Klink, R. R., & Guo, J. (2013). Creating gender brand personality with brand names: The effects of phonetic symbolism. *Journal of Marketing Theory and Practice*, *21*(3), 319–330. https://doi.org/10.2753/MTP1069-6679210306

Zaltzman, H. (2023, January 13). *Debuts* (Vol. 168) [Podcast]. https://www.theallusionist.org/transcripts/tag/snipa