DOI: 10.1111/phpr.13110

ORIGINAL ARTICLE

Philosophy and
Phenomenological Research

What is conceptual engineering good for? The argument from nameability

Steffen Koch[1] | **Gary Lupyan**[2]

[1]Abteilung Philosophie, Universität Bielefeld, Bielefeld, Germany

[2]Department of Psychology, University of Wisconsin-Madison, Madison, Wisconsin, USA

Correspondence
Steffen Koch, Universität Bielefeld, Abteilung Philosophie, Postfach 100131, D-33501, Germany.
Email: steffen.koch@uni-bielefeld.de

Funding information
This work was supported by NSF PAC awarded to G. Lupyan, Division of Behavioral and Cognitive Sciences, Grant/Award Number: 2020969

Abstract
It is often assumed that how we talk about the world matters a great deal. This is one reason why conceptual engineers seek to improve our linguistic practices by advocating novel uses of our words, or by inventing new ones altogether. A core idea shared by conceptual engineers is that by changing our language in this way, we can reap all sorts of cognitive and practical benefits, such as improving our theorizing, combating hermeneutical injustice, or promoting social emancipation. But how do changes at the linguistic level translate into any of these worthwhile benefits? In this paper, we propose the nameability account as a novel answer to this question. More specifically, we argue that what linguistic resources are readily available to us directly affects our cognitive performance on various categorization-related tasks. Consequently, our performance on such tasks can be improved by making controlled changes to our linguistic resources. We argue that this account supports and extends recent motivations for conceptual engineering, as categorization plays an important role in both theoretical and practical contexts.

KEYWORDS
categorization, conceptual engineering, linguistic interventions, nameability

1 | INTRODUCTION

It is often assumed, both in academia and in public discourse, that how we talk about the world matters a great deal. This is one reason why it is so controversial how we talk about issues of political importance, such as sex and gender, race and racism, public health, climate change, taxation, and sexual harassment. In philosophy and related disciplines, the assumption that language is an artifact that requires our constant attention and maintenance is what drives projects in so-called "conceptual engineering" (or "ameliorative projects"). In the words of Herman Cappelen, conceptual engineering is "the process of assessing and improving our representational devices" (Cappelen, 2018, p. 3), where "representational devices" can include words, concepts, and possibly other things. Much of the current debate focuses on conceptual engineering about language, which will also be our target.[1]

Language-centered conceptual engineering can take different forms. On the one hand, there are proposals to introduce, eliminate, replace or revise our usage (or the meaning) of *specific expressions*; for example, to stop using expressions like "fake news" (Habgood-Coote, 2019), "democracy" (Cappelen, 2023), or "concepts" (Machery, 2009); to replace the term "social distancing" with "physical distancing" (Schnell 2020); or to revise our use of terms like "woman" or "man" (Haslanger, 2012), "race" (ibid.), "misogyny" (Manne, 2017), and "sexual orientation" (Dembroff, 2016). On the other hand, there are proposals to change certain *structural features* of a given language; for example, to eliminate gender-specific terms from the English language (Dembroff & Wodak, 2021) or to not make use of generic generalizations in the social realm (Leslie, 2017; see also Wodak et al., 2015). We will mainly be concerned with the former type of conceptual engineering, but our assessment of it will have some more structural implications as well.

Conceptual engineers typically assume that by changing our language we can reap all sorts of cognitive and practical benefits. Some of them are epistemic: We can develop better theories, acquire more consistent world views, or increase the accuracy of our inferential patterns (Fischer, 2020). Others are social or moral: We can overcome hermeneutic injustices that arise from unjust gaps in the conceptual resources needed to express ourselves (Fricker, 2007), acquire linguistic resources that help us to better navigate the social world (Machery, 2017, pp. 216–217), or contribute to social emancipation (Haslanger, 2012). As Ritchie puts it, "all ameliorative projects [are] aimed at changing how we talk and think as a partial means to change how we behave" (Ritchie, 2021, p. 469). Or, as Riggs puts it, conceptual engineers "don't want to change meaning for its own sake," but see it as an appropriate means "to change some other thing, e.g., our social relations or our philosophical theorizing" (Riggs, 2019, p. 16).

It is far from clear, however, how changes at the linguistic level translate into any of these worthwhile benefits. Most of the current literature on conceptual engineering seems to take it for granted that conceptual engineering has, or at least can have, such worthwhile benefits, without providing a detailed analysis of how this might be. What *does* seem clear is that if conceptual

[1] See Isaac et al. (2022) and Koch et al. (2023) for recent literature reviews on conceptual engineering.

engineering is to change our theorizing and social relations for the better, it will have to take a cognitive detour. Language change does not have an immediate effect on our theorizing or our social relations. Rather, language change is associated with cognitive changes that, ostensibly, enable speakers to theorize better or lead to new forms of behavior. But what exactly are the cognitive changes that can be expected to be induced by conceptual engineering?

Some philosophers assume a strong connection between linguistic resources on the one hand and our ability to think and talk about a phenomenon on the other. In this vein, Sterken writes that adding or revising 'word-meaning pairs' to a language through conceptual engineering may

> "enable an interpreter to think and communicate things she could not have thought or said without having that meaning – having that meaning gives the interpreter new abilities to imagine, recognize, create cognitive models, and communicate using that meaning." (Sterken, 2020, p. 430)

Sterken also notes a connection between the lexicon of a language and the ability of speakers to recognize a phenomenon. Using the example of 'post-partum depression', she claims that "if our language lacks the word-meaning pair of 'postpartum depression'," then the sufferings of young mothers "might go unrecognized" (ibid., p. 418).

As we will argue later on, there is something true in Sterken's remarks; but it is not at all trivial to spell out *just how* lexicalizing some phenomenon contributes to our ability to recognize, think, or even talk about it.[2] After all, we can easily think and talk about very soft blue sofas with yellow dots, desks with wooden tops and steel legs, and prime numbers between 1,000 and 2,000 that contain the number 7 without having conventional terms for them. We are also able to recognize these things when we see them. The same, one might argue, should hold for 'postpartum depression' and 'sexual harassment'.

So is the value of conceptual engineering simply to provide us with handy labels to stand in for what would otherwise be complex descriptions? Is it about what Deutsch calls "syntactic convenience" - "to replace longer descriptions [...] with a shorter, single term" (Deutsch, 2020, p. 3945)? Syntactic convenience is certainly not unimportant. It makes our exchanges more effective, reduces the cognitive load during language processing, and helps us remain within the word limits of our favorite academic journals. Nonetheless, we agree with Deutsch that if syntactic convenience were *all* we could gain from conceptual engineering, this would be bad news for self-styled conceptual engineers. It would make it difficult to see how conceptual engineering could be an appropriate means of achieving theoretical benefits or improvements in our social relations.

All of this raises the challenge for conceptual engineers to find a rationale for their practice that combines two important features: (i) it is supported by, or at least consistent with, the available empirical evidence about the interface between language and thought; and (ii) it explains why conceptual engineering has cognitive benefits that go beyond mere syntactic convenience and that relate to the goals of conceptual engineers. The main goal of this paper is to develop an account of the cognitive effects of conceptual engineering that meets this challenge. We shall call it the *nameability account*. As we will show, the crucial ingredients of this account are supported by robust empirical evidence. In addition, the account demonstrates that conceptual engineering

[2] Here and elsewhere, we will say that something is "lexicalized" in a language if it is expressed through a single, non-compositional lexical unit (cf. Rissmann et al., 2023).

has interesting cognitive effects that go beyond mere syntactic convenience and that are plausibly related to the goals formulated by conceptual engineers.[3]

The remainder of the paper is organized as follows. In Sect. 2, we present a general outline of the account, identify its basic ingredients, and explain its advantages. In Sects. 3–5, we develop and defend each of the account's three ingredients in detail. In Sect. 6, we discuss some of the most pressing objections to the account. Finally, in Sect. 7, we draw some broader implications from the account, also relating it to some typical goals of conceptual engineers.

2 | THE NAMEABILITY ACCOUNT

The account we develop in the following pages contains three crucial ingredients. In this section, we will give an overview of the account and explain some of its benefits. In the next sections, we will develop each of its ingredients in more detail.

To state our account precisely, let *C* be some category of entities (objects, events, properties, relations); let categorization be the practice of judging whether some entity *e* instantiates a given category or not; let *c*-categorization be the practice of judging whether *e* instantiates *C* or not; and let *G* be a group of people. Now the account can be summarized thus:

The Nameability Account

(i) Categorization
For many groups *G* and categories *C*, increasing the accuracy of *G*'s c-categorization has significant non-communicative value.

(ii) Nameability
The accuracy of *G*'s c-categorization increases if the nameability of features that are diagnostic of *C* increases within *G*.

(iii) Conceptual engineering
Conceptual engineering can increase the nameability of features that are diagnostic of *C* within *G*.[4]

Note that these three ingredients can serve as premises in an argument for the value of conceptual engineering. If we accept all of them, it directly follows that conceptual engineering can have significant non-communicative value, that is, value manifested in non-communicative situations.

The notions of nameability and diagnostic features need spelling out. We will give detailed accounts of both in subsequent sections. For now, the following rough understanding shall suffice: Nameability concerns how easy it is for people to name something, that is, to produce an adequate verbal description of it. A diagnostic feature is one that people use in deciding whether or not something *e* instantiates a category *C* (this may or may not be *C*-ness itself). So, *Nameability* says that if the features that people use in categorizing something become easier for them to name,

[3] Nameability effects of the sort under consideration have so far been neglected in philosophy – at least in the literature on conceptual engineering that we are considering here. One notable exception is Ritchie (forthcoming), who argues that nameability effects may play a role in the context of labeling previously unlabeled social identities (pp. 10–11).

[4] In what follows, we will use italicized forms of "categorization", "nameability" and "conceptual engineering" to refer to the claims listed here.

then they will become more accurate in their categorization practice. We will provide empirical evidence for this claim in sect. 4.

Before we go into more detail, we wish to spell out some of the benefits of the nameability account of the value of conceptual engineering. First, the account is *domain-general*. There is no principled restriction to what can be inserted for *C*. As we shall see later on, the account thus establishes the value of conceptual engineering for both epistemic/scientific and ethical/societal purposes. Note also that the account is not restricted to either *de novo engineering* or *conceptual re-engineering* (Chalmers, 2020). In principle, a feature can be made more nameable by introducing a new term to denote it, or by revising an existing one.[5]

Second, all of the crucial ingredients that the nameability account uses – categories, how accurate people are at categorizing things, and how nameable a given feature is – are *empirically tractable*. There are established empirical paradigms for measuring people's accuracy in categorization and for measuring the nameability of a feature. Crucially, the argument does not rely on any potentially contentious notion of concepts to establish its conclusion. To be sure, one might want to invoke concepts to explain *why* nameability affects categorization; but the claim that it *does* is directly supported by empirical evidence and thus not dependent on any particular view of concepts or categorization.

Lastly, the account establishes a *causal role* for conceptual engineering in the production of non-communicative value. It makes explicit how engaging in conceptual engineering sets in motion a causal chain that may result in positive worldly effects that go beyond communication.

With these remarks in the background, let us now move on to a more detailed defense of *Categorization, Nameability*, and *Conceptual engineering*. We begin with *Categorization*.

3 | CATEGORIZATION

This section serves to flesh out and defend the first element of the nameability account: the claim that, for many groups *G* and categories *C*, increasing the accuracy of *G*'s c-categorization has significant non-communicative value.

Categories, as we understand them here, are collections of things (objects, events, or properties) in the world (Murphy, 2002, p. 1; Rosch, 1978, p. 28): scissors, cell phones, inauguration ceremonies, weddings, or basketball games. Categories must be distinguished from our mental representations (or concepts) of them. This, however, does not mean that categories cannot be (partly or wholly) constituted by our social practices, norms, and representations. One way to see this is that there are cross-cultural and cross-temporal differences in our categorizing practices. What is considered polite in one part of the world may not be so in another; and we have all experienced that what was once fashionable is no longer (though it may be again in a few years). Other categories are assumed to exist independently of our actions. On a popular line of reasoning in philosophy, natural kinds like water, tigers, gold, or elm trees are joints in the natural world, waiting to be discovered rather than being invented by us.

Even if a category depends on cultural norms and practices, however, it does not mean there is no way to compute categorization accuracy. Individuals can be wrong about whether a given object belongs to the category of scissors, laptops, or lamps, or whether a given behavior instantiates a wedding, a basketball game, or an arrest, to the extent that their categorization deviates from

[5] Which of these two options is the better one in a given case is a complicated question that stands orthogonal to the goals of this paper. See Koch (2024) for discussion.

a given (group) norm. Importantly, this holds even for *ad hoc* categories, such as those invented by an experimenter.

Categorization is the process by which we group the things we encounter into categories. Croft and Cruse describe categorization as follows:

> Categorization involves the apprehension of some individual entity, some particular of experience, as an instance of something conceived more abstractly that also encompasses other actual and potential instantiations. For instance, a specific animal can be construed as an instantiation of the species DOG, a specific patch of color as a manifestation of the property RED, and so on. (Croft & Cruse, 2012, p. 74)

When we categorize an entity, we subsume it under a type. This, in turn, generates "inference tickets" of the following form:

(i) *e* is *F*.
(ii) *F*s are (very likely to be) *G*.
(iii) Therefore, *e* is (very likely to be) *G*.
(or, reversely: *e* is not an F; only *F*s are *G*; therefore, *e* is not *G*.)

Categorization is an (often automatic) process that is not immune to error. Categorization errors come in two basic forms. We can *false alarm*: categorize an entity *e* as belonging to a category *C*, even though *e* is not an instance of *C*. For some categories, this may be because it is an objective fact that *e* is not *C*; for others, it may be because categorizing *e* as *C* deviates from a relevant norm. We can also *miss*: fail to categorize *e* as *C*, even though *e* is *C*. Again, this general idea permits that there may sometimes be a worldly (non-social) explanation for the fact that *e* is *C* and sometimes just a social one. Accordingly, there are two ways to improve categorization accuracy. The first is to lower false alarms, i.e., misclassifying fewer things as *C*s that aren't *C*s. This increases our *specificity*. Specificity is defined as the probability of not judging that *e* is *C*, given that *e* is not *C*. The other is if we miss fewer instances of *C*. This increases our *sensitivity*, which is defined as the probability of judging that *e* is *C*, conditional on *e* being *C*. *Accuracy* combines sensitivity and specificity. Perfect accuracy requires maximal specificity and maximal sensitivity.[6]

Why does accuracy matter? The short answer is: because both false alarms and misses can be very costly. Consider misses first. From an epistemic perspective, not noticing that *e* belongs to *C* involves a loss of knowledge. If you don't *notice* that *e* is *C*, you won't come to *know* that *e* is *C*. Moreover, not noticing that *e* is *C* prevents you from using the inference tickets from *C*-hood to properties that are highly correlated with *C*-hood, such as *F*-ness or *G*-ness. These epistemic costs easily generate practical costs as well. If you don't see that *e* is *C*, then you won't respond to *e* in a way that is appropriate to *C*. This can have multiple consequences. If *C* is something dangerous, then failing to categorize *e* as *C* can lead to risky behavior (suppose *e* is a bomb or a deadly infectious disease, and you fail to categorize it as such). If *C* is something good, it can lead to missing an opportunity (suppose *e* is your favorite berry and you end up not eating it). If *C* is a social group whose members want to be treated in a certain way, then failing to categorize *e* as a member of *C* can lead to inappropriate behavior toward *e* that causes harm and is socially

[6] Specificity and sensitivity can be individually maximized in a trivial way. We become maximally specific if we stop classifying anything as *C*, and we become maximally sensitive if we classify everything as *C*. Ordinarily, we need to find the right balance between specificity and sensitivity depending on whether misses or false alarms are more costly.

sanctioned. In sum, correctly categorizing entities, in the sense of avoiding misses, is important both for epistemic reasons and because categorization is linked to behavior.

Now consider the costs of false alarms. From an epistemic perspective, false alarms lead to any number of false beliefs: the belief that *e* is *C*, and many of the beliefs that follow from that belief, e.g., that *e* has a property *F* that is typical of members of *C*. This also has practical consequences. If you mistakenly believe that *e* is *C*, then you will respond to *e* in a way that is appropriate for *C*, but not necessarily for *e*. Again, this can lead to risky behavior or missed opportunities. For example, suppose you mistakenly categorize *e* as edible. You are then likely to eat *e*, which could result in health risks or social sanctions. Or suppose you mistakenly categorize a fabric as heat resistant. You might then use this fabric to touch hot goods and burn yourself. Or suppose *C* is a social kind to which *e* does not belong. Then falsely categorizing *e* as *C* might lead to inappropriate behavior toward *e*. It seems, then, that avoiding false alarms is important for pretty much the same reasons as avoiding misses.

If the levels of specificity and sensitivity in our categorization practice have important practical and epistemic downstream effects, then so does accuracy. This has the direct implication that increasing the level of accuracy of a given categorization practice can have very important worldly consequences, at least if the category in question plays a prominent role in our lives. The positive effects that an increase in accuracy can have range from epistemic benefits, such as avoiding false beliefs or gaining knowledge or understanding, to practical issues, such as showing appropriate responses to dangers, opportunities, or other people, or simply being successful in one's actions.

None of this implies that all categories are equally important. Categories can be irrelevant or even harmful for all sorts of reasons. Consider David Lewis's invented category consisting of the right half of my left shoe, the moon, and the sum of all of Her Majesty's earrings (Lewis, 1986, p. 213). This category is completely useless (except when arguing about mereological composition!). Whether a category representation is helpful, irrelevant or detrimental, typically depends on your broader interests. Nonetheless, there are *some* categories such that it would have positive worldly consequences if *at least many* people were more accurate at categorizing things according to them.

4 | NAMEABILITY

Now that we have elaborated on the first element of the nameability account, *Categorization*, it is time to move on to the second, *Nameability*. This is the claim that the accuracy of *G*'s c-categorization increases if the nameability of features that are diagnostic of *C* increases within *G*.

Nameability posits a connection between two measurable quantities: how accurate people are at categorizing things as *C*, on the one hand, and how nameable the features they rely on to categorize them – so-called diagnostic features – are, on the other. Let us first explain nameability and diagnostic features, before outlining the empirical evidence for the link between nameability and categorization.

Nameability (not in italics) refers to how easy it is for people to produce an adequate verbal description of something (an object, an event, a property, or a relation) (Lupyan & Zettersten, 2021, p. 174). How nameable something is varies across individuals and groups, depending on people's knowledge of the object domain, their cultural background, and, most importantly, the languages and idiolects they speak. As a result, something that is highly nameable for you may be less nameable for people on the other side of the globe, and *vice versa*. For typical American English speakers things like horses, cars, or baseball games are highly nameable, but things like chevrotains, thunderbolts, or corroborees are not. Aboriginal Australians, on the other hand, will have no trouble naming a corroboree when they see one, just as people familiar with ancient Greek mythology won't have difficulty naming thunderbolts.

Nameability can be operationalized in at least two ways. The first is *naming coherence*: how coherent people are in how they respond to a naming task, where coherence is measured by the uniformity of responses across different instances. High coherence corresponds to high nameability. Naming coherence can be measured at the individual or at the group level. Individual-level naming coherence concerns diachronic naming coherence within *a single* individual; group-level naming coherence concerns (synchronous or diachronic) naming coherence between *different* individuals. Interestingly, individual-level and group-level naming coherence are highly correlated (ibid.), which allows using measures of group-level naming coherence to predict individual-level naming coherence.[7]

A second way to operationalize nameability is through *naming complexity*: how many words or clauses people need to name something. The fewer words or clauses the average person needs, the more nameable the thing in question. An important factor in naming complexity is, of course, whether the thing in question is lexicalized in the language spoken by the members of that group. If a given language marks an object or distinction with a compact label, then speakers of that language will typically need fewer words to describe the thing in question, at least if they are familiar with that label. Although there is no logical connection between naming coherence and naming complexity, in practice the two measures are highly correlated. In the following, both naming coherence and naming complexity will play a role.

Next, we need to explain what it takes for a feature to be *diagnostic* of a category. The idea is quite straightforward: When we classify entities into categories, we rely on certain easily recognizable features that are highly typical of members of the given category, such that the possession of these features correlates with category membership. As Tversky and Kahnemann put it, an attribute or feature is highly diagnostic of a class "if the relative frequency of this attribute is much higher in that class than in a relevant reference class" (Tversky & Kahnemann, 1983, p. 296). The diagnosticity of a feature F relative to a category G can be formally described as the ratio of the probability that e has Ff conditional on e being G to the probability that e has F conditional on e not being G (cf. Gennaioli & Shleifer, 2010). Although this formula is not defined for such a case, it is important to note that all categories are maximally diagnostic of themselves. There is a perfect correlation between something being a dog and it belonging to the category of dogs.

Nameability states that there is a connection between the nameability of the diagnostic features of a category and how accurate we are at categorizing things according to that category. There is accumulating evidence that this connection holds. Instead of offering an exhaustive review of the relevant empirical work, we will concentrate on three particularly salient experiments. In all of them, nameability appears to be a causal factor in how people perform in non-linguistic categorization tasks. What makes the results of these experiments especially remarkable is that they are conducted with very concrete, visually discriminable categories such as shapes and colors. If nameability has a proven effect even on how we perform with *those* categories, it can be expected to have even larger effects when it comes to more abstract categories such as sexual harassment or misogyny (more on this in sect. 6).

[7] The general strategy for measuring group-level naming coherence is simple: present a stimulus to multiple participants and ask them to say what it is. There are several ways to calculate naming coherence from such data, ranging from consistency-based measures (e.g., Brandimonte et al., 1992), over entropy-based measures (e.g., Brodeur et al., 2010) to divergence-based measures. The differences between these measures are rather subtle and need not detain us here.

Nameability effects on problem solving: Bongard problems Bongard problems are a type of categorization task that consist of 12 pictures: 6 on the left and 6 on the right. The challenge is to find the feature that distinguishes the 6 images on one side from the 6 images on the other. Lupyan and Zettersten (2021) provide evidence that the nameability of the distinguishing feature is a strong predictor of success in solving a Bongard problem. Analyzing data from Foundalis (2006), they found that participants' success in solving Bongard problems was highly correlated with the complexity of the verbal descriptions that the inventors of the problems used to formulate the correct solutions. This hypothesis was then tested in a controlled experiment in which participants were asked to state their verbal answers to Bongard problems. The experimenters found that problems whose correct solutions exhibited higher average verbal complexity and lower naming coherence (measured by the percentage of unique words used across correct responses) were also more difficult for people to solve. Nameability ends up being a strong predictor of how accurately people detected the target category, i.e., the solution to a given Bongard problem.[8]

Nameability effects on similarity judgments: shapes and surfaces Nameability effects also appear in another type of categorization task. Lupyan and Zettersten (2021) showed participants series of shapes and asked whether one of the shapes (*a*) was more similar to shape *b* or shape *c*. Shapes *a* and *b* had the same global contour while shapes *a* and *c* had the same surface structure, e.g., smooth, curved, bubbly, fluffy, etc. Overall, participants showed a strong bias to match based on the global contour. However, the likelihood that this strong bias was overridden by the surface structure was strongly predicted by nameability: the more nameable structures were used much more as the basis of grouping than the harder to name surfaces. The nameability of a given surface structure was measured in a separate test in which participants were asked to describe the surfaces in question. Their nameability was then calculated using the measures described above. It turned out that more nameable surface outlines were more likely to be chosen as the relevant feature by which to group the items. The implication of these results is that as things become more nameable, they become more obvious dimensions to use as a basis for categorization. And although in this particular experiment both choices were "correct," this implication also applies to contexts in which only one of the available choices is correct. In such contexts, increasing the nameability of a relevant category makes people more accurate at categorizing things according to it.

Nameability effects in category-learning: colors In a study by Zettersten and Lupyan (2020), the experimenters sought to test whether nameability also affects how easy it is for people to learn novel rule-based categories. They proceeded in two steps. First, using data from a large-scale online color naming study (N = 134,727; Munro, 2010), they determined the nameability of a large number of different colors. Next, they created sets of highly nameable colors (80-85% group-level naming coherence) and sets of poorly nameable colors (6-10% group-level naming coherence), holding the discriminability between colors fixed. Each of these sets was then used to construct categories with identical structures. For each, a single color was perfectly predictive (or diagnostic) of category membership. It turned out that participants were more accurate in categorizing the color plates when the diagnostic features were highly nameable colors, such as red and brown, than when they were colors with low nameability, such as lavender and olive. They were also faster to learn the given category. These results were obtained in both between-subjects and within-subject designs, and they were replicated when shapes were used instead of colors. Thus,

[8] Lupyan and Zettersten (2021) further replicated these results using simplified partial Bongard problems they developed on their own and tested them with more participants. Again, they found a correlation between average accuracy and verbal complexity.

participants were more accurate at categorizing the color plates according to a pre-given category (invented by the experimenters) if the diagnostic features of this category were more nameable.[9]

The evidence reviewed above suggests the following: All other things being equal, the nameability of a feature affects how likely we are to use that feature even in contexts not requiring explicit naming. If this feature is indeed sufficiently diagnostic of the category we are trying (or are supposed) to track with our categorization practice, then it may significantly improve our performance in that practice. This has the following consequence: If a feature *F* is sufficiently diagnostic of membership in a category *C*, then increasing the nameability of *F* will increase our accuracy in categorizing things as *C*.[10]

Although language is only one of potentially several levers that can be pulled to increase our accuracy in categorizing things, recent research suggests that its influence may be special. One reason is that while other ways of activating a category representation, such as sounds or appearances, covary with the properties of the event that produced them, words do not (Edmiston & Lupyan, 2015). For example, a photograph of a dog will never show just a dog: it will show a particular dog, of a particular breed, in a particular pose. The same is true of sounds such as dog barks. This makes words a particularly efficient means of activating basic-level category representations (Murphy, 2023; Rosch, 1978), that is, representations of *tigers*, *dogs*, or *monkeys*, rather than representations of, say, *Fido the dog* or *male chimpanzees napping in a tree*. This suggests that modulating the nameability of a feature through conceptual engineering is a particularly effective means of increasing the accuracy of related categorization practices.

Our main focus in this section has been to draw connections between nameability and categorization accuracy in contexts that do *not* require explicit naming. It is worth noting, however, that nameability also has demonstrated effects on communication. Rissman et al. (2023) investigated the consequences of having or not having a conventional term on a speaker's ability to communicate superordinate categories such as beverages, vehicles, or appetizers. To do this, they had English and Mandarin Chinese speaking adults play a director/matcher communication game. On each trial, a "director" saw a 3-by-3 word grid with three words highlighted. The "matcher" saw a grid with the same words, but no highlighting. The director's task was to give the matcher a verbal cue that would allow the matcher to select those three (and only those three) words. The authors found that participants were more successful if they could rely on a conventional superordinate term rather than a compositional description, even in cases when it would *seem* like the superordinate term could easily be paraphrased. For example, Chinese has a word that denotes motor vehicles (chēliàng) and a word that denotes boats (chuán), but not a single term that includes both. A Chinese-speaking director faced with giving a clue that would allow a Chinese-speaking matcher to select "airplane", "tractor", and "boat", *could* use a conjunction of chēliàng and chuán (and many did); nevertheless, English speakers who could rely on the conventional superordinate term "vehicle" performed better. Other things being equal, more nameable categories seem to allow for smoother and ultimately more successful communication.

[9] While the experiments described here were all run with adults, Zettersten et al. (2023) found similar but smaller effects with 4-6 year old children.

[10] Other empirical work points in a similar direction. Teaching people new verbal labels directly affects their non-linguistic categorization practices (Christe & Gentner, 2013; Lupyan & Casasanto, 2015). There is also evidence that: named categories are easier to learn (Balaban & Waxman, 1997; Lupyan, Rakison, & McClelland, 2007; Nazzi & Gopnik, 2001; Perry & Samuelson, 2013; Plunkett, Hu, & Cohen, 2008), category knowledge is more effectively activated by verbal than by non-verbal cues (Boutonnet & Lupyan, 2015; Lupyan & Thompson-Schill, 2012), naming impairments such as aphasia lead to categorization impairments (Gainotti, 2014; Lupyan & Mirman, 2013 for review), and interfering with language in healthy adults impairs categorization (Lupyan, 2009).

5 | CONCEPTUAL ENGINEERING

In the previous two sections, we have made a case for the first two elements of the nameability account by showing why accuracy in our categorization practices matters and how nameability contributes to it. What is left for us to show is how conceptual engineering may contribute to making things more nameable, that is, the claim we earlier labeled *Conceptual Engineering*.

Nameability, as characterized above, has two separate but correlated dimensions – naming complexity: the complexity of the verbal description needed to pick out a given thing, and naming coherence: the degree to which individuals and groups show consistency in how they verbally respond to a thing. The experiments described above show that both of these dimensions predict how well we do at categorizing things. This suggests that conceptual engineering, when applied to appropriate features and categories, could have a positive impact if it could help reduce naming complexity, or if it could help foster naming coherence across a population.

Both of these things – reducing naming complexity and fostering naming coherence – are part of the conceptual engineer's job description. First, consider how conceptual engineering serves to reduce naming complexity. De novo engineering is about inventing new linguistic tools to talk about features or categories that were not previously lexicalized in a language. By inventing such tools and disseminating them in a linguistic community, it increases the nameability of the feature or category in question. To illustrate, before the term "sexual harassment" was introduced, sexual harassment occurred, and it was *possible* to think and talk about it in one way or another. However, the lexicalization of the phenomenon drastically reduced the verbal complexity needed to describe the phenomenon, thereby making it more nameable. Similarly, while it was possible to talk and think about the phenomena now called "mansplaining" and "postpartum depression," the invention of the labels served to make them more nameable.

Conceptual re-engineering, that is, changing the conventionalized use or meaning of a pre-existing term, can have similar effects. If we already have a compact verbal label V, and we then change our way of speaking so that we now apply V to a previously unlabeled feature F, then F becomes more nameable through this process. For example, when Manne defines "misogyny" as roughly *the hostile social forces girls and women face that serve to police and enforce a patriarchal order* (Manne, 2017), she makes this phenomenon more nameable. Those of us who read and adopt her proposal will henceforth be equipped with a compact label to refer to this phenomenon.[11]

But conceptual engineering is not just about introducing new verbal labels or proposing changes in the use of existing ones. It is also about taking steps to implement such proposals, that is, to spread a newly introduced verbal label, or a new way of using an existing one, throughout a linguistic community. If successful, this second aspect of conceptual engineering serves to promote naming coherence – to reduce the divergence in how different people refer to phenomena of interest.[12]

[11] We note that the nameability account could also be used to support the cognitive value of *elimination* rather than introduction or re-engineering. The evidence reviewed above suggests that making a feature *less* nameable by eliminating the default expression for it has the effect of making people less likely to use that feature in categorization. For certain categories and groups, this might be just what we want.

[12] Note, however, that implementing proposals to *change* a given linguistic practice, e.g., by using an entrenched word in a novel way, might first lead to a reduction in naming coherence, because the new usage will compete with the entrenched one. An increase in naming coherence is thus contingent upon being *successful* in one's implementation strategy.

There is currently some debate about whether and how implementation can be achieved. Some philosophers have argued that conceptual engineering can be implemented by manipulating the social norms that govern our linguistic behavior (Löhr & Michel, 2023; Nimtz, 2021). Following Bicchieri (2016), these philosophers understand social norms as rules of behavior to which people adhere because they have two expectations: (a) that most people in their reference network – the people to whom they adhere in relevant matters – do in fact adhere to these rules (empirical expectation), and (b) that most people in their reference network believe that they should adhere to them (normative expectation). Given this, social norms and, *a fortiori*, our linguistic behavior can be changed either by shifting people's reference network or by convincing members of that network to publicly adopt new and different ways of using terminology.

There is no guarantee that even well-supported conceptual engineering projects will be successful. The nameability account does not imply a high success rate for such projects. But the account gives a clear indication of the effects that successfully implemented conceptual engineering projects give rise to. Standardized ways of referring to phenomena across a linguistic community increase their nameability, thereby increasing the accuracy with which speakers categorize them and other categories that they are diagnostic of.

6 | OBJECTIONS

For the nameability account to be successful, it must satisfy the following two conditions: (i) nameability must be *causally* related to the accuracy of a categorization practice (mere correlation is not enough) and (ii) nameability effects on categorization must occur for the kinds of categories that conceptual engineers typically target. In this section, we discuss possible objections to both of these conditions.

6.1 | Causation vs correlation

The question with which this paper began was whether conceptual engineering can be used in the service of producing non-communicative benefits. To do so, it is not enough that increased nameability correlates with categorization accuracy; there must be a (direct or mediated) causal relationship between the two, such that increased nameability reliably leads to increased categorization accuracy. Looking at the experiments described above, one might wonder whether it is really nameability that is in the driver's seat, or whether nameability is merely piggybacking on something that might not be present in the cases of interest to conceptual engineers. Here we will focus on familiarity and memorability as possible confounds.

That familiarity, rather than nameability, drives the results may seem plausible for the Bongard problem experiment. After all, not only are shapes like triangles more nameable than most other shapes, but the average adult living in the US or Europe may also be more familiar with things like triangles, squares or circles than with many other shapes. However, even if this alternative explanation is plausible for Bongard problems, it is unlikely to apply to other experiments that elicit nameability effects. For example, analyses of color distributions in natural and urban scenes show that both are dominated by low-saturation colors, which tend to be less nameable than high-saturation colors (Belpaeme & Bleys, 2009; Yendrikhovskij, 2001; Zettersten & Lupyan, 2020). Thus, there is little reason to expect that participants in the high nameability condition of the color experiment had the benefit of more frequent exposure to (and thus familiarity with) the colors they were presented with. Moreover, the shape and surface structure experiments seem immune to this explanation, as none of the shapes used in these experiments were particularly familiar to participants.

Memorability may indeed play a role in categorization accuracy. If a feature is more memorable, you are more likely to think about it in subsequent categorization tasks; and if this feature is indeed highly diagnostic of a given category *C*, then this directly affects your performance in c-categorization. But as Zettersten and Lupyan point out, memorability is a confound only if it causes differences in categorization accuracy independently of naming (Zettersten & Lupyan, 2020, p. 19). For color, however, studies suggest that the opposite is true: memory patterns are best predicted by naming patterns, not the other way around (Davidoff et al., 1999; Roberson et al., 2000). To the extent that naming and memory are indeed related, the causal arrow seems to run from name-based categorization to memory rather than the other way around. This means that, at least with respect to the color study, memorability is unlikely to be a confound. All in all, the diversity of the experimental designs and category domains used to test nameability effects makes it unlikely that there is a common confound to the experiments discussed above.

6.2 | Concrete vs. abstract categories

The categories tested in the experiments discussed above – geometric figures, colors, and shapes with particular surfaces – are all fairly concrete. Compare these categories with those typically targeted by conceptual engineers, such as sex, gender, race, misogyny, knowledge, belief, or planet. These categories seem more abstract and less susceptible to categorization on purely perceptual grounds. The distinction between representations of concrete and abstract categories is controversial among philosophers and psychologists. Many theorists in the field believe that there is a psychologically real difference between our representations of concrete and abstract categories, e.g. in terms of recognition (Strain et al., 1995), acquisition (Schwanenflugel, 1992), or stored information type (Barsalou & Wiemer-Hastings, 2005). Now, if (i) there is a psychologically significant distinction between our representations of concrete and abstract categories, (ii) nameability effects have only been found for concrete categories, and (iii) conceptual engineers typically target abstract categories, then it's not clear that the nameability account actually supports conceptual engineering.

Nameability studies with more abstract categories would indeed be helpful in determining the scope of the nameability account. But even in the absence of such studies, there is reason to think that nameability effects should, if anything, be *larger* for abstract categories. The reason is that many abstract categories depend more on language for being identified than concrete ones (indeed, the need to rely on language to communicate the category was how abstractness was defined by Brysbaert and colleagues when they collected their widely-used lexical concreteness norms; Brysbaert et al., 2014; cf. Langland-Hassan & Davis, 2023). While nameability can quantitatively increase the accuracy of a given categorization practice (as the experiments discussed above show), many abstract categories may not be identifiable without language at all (e.g., see Borghi, 2020 for discussion). This suggests that our performance with abstract categories is *more*, rather than *less*, dependent on nameability. The experiments reviewed above show that there are nameability effects even for categories that can be identified by non-linguistic means (such as perception). Why should these effects disappear in cases where we do not have these alternative means (or have them to a lesser extent)?

7 | THE NAMEABILITY ACCOUNT AND THE GOALS OF CONCEPTUAL ENGINEERING

Let us take stock. We have now defended each of the three elements of the nameability account. First, we have argued that, at least for many groups *G* and categories *C*, increasing the accuracy of *G*'s c-categorization brings important non-communicative benefits. Second, we have shown that the nameability of the features that are diagnostic of *C* (including *C* itself) is a strong predictor of people's c-categorization performance. Finally, we have argued that, more or less by definition, successful instances of conceptual engineering serve to make relevant features more nameable. Taken together, this shows that conceptual engineering can increase the accuracy of people's categorization practices, and thus may bring important non-communicative benefits.

What else does the nameability account teach us about conceptual engineering? The first thing to note here is that it provides some justification for Sterken's postulated link between the lexicalization of a category and the recognition of that category. Other things being equal, making something more nameable in a linguistic community – for example, by introducing a compact label to refer to it – increases the likelihood that members of that community will pay attention to it and use it in their categorization practice. The introduction of the label "postpartum depression" and the spread of its use within a community likely increases public awareness of the depressive condition that affects up to 15% of new mothers after childbirth. The word not only helps communicate about postpartum depression after it is diagnosed – it increases the likelihood that it will be diagnosed in the first place.[13]

A second thing worth mentioning concerns Deutsch's criticism that conceptual engineers have little more to offer than syntactic convenience. In one way of putting the results of the discussion above, this is clearly wrong. As we have seen, conceptual engineering can also increase the accuracy of people's categorization practices and make communicative exchanges more successful (as evidenced in the Rissman et al. study discussed at the end of sect. 4). Another way of putting this point is to say that Deutsch is quite right that the primary effect of conceptual engineering is syntactic convenience – he just underestimates the communicative and non-communicative effects of syntactic convenience. For, as we have seen, having a short label to refer to something has important cognitive effects beyond mere convenience: it affects how we perform in relevant categorization practices and how successful we are in getting others to understand our communicative intentions.

How does the nameability account relate to the goals formulated by conceptual engineers? As we have argued, conceptual engineering affects our categorization practices, which in turn affect our behavior. Thus, in general, Ritchie's claim that conceptual engineering can be "a partial means to change how we behave" (Ritchie, 2021, p. 469) is well in line with the nameability account. We also believe that the effects of increased nameability are directly related to other, more concrete benefits described by conceptual engineers.

One commonly stated goal of conceptual engineering is to increase the accuracy of our inferential patterns (Fischer, 2020). As we pointed out in Sect. 3, categorization is linked to inference by licensing inferences of the form: (i) *e* is *F*; (ii) *F*s are (very likely to be) *G*; (iii) therefore *e* is (very likely to be) *G* (or, conversely, *e* is not an *F*; only *F*s are *G*; therefore *e* is not *G*). But these inferences are only as accurate as their first premises. By becoming more accurate in categorizing all and only those things under a category that actually belong there, we can increase the

[13] Making postpartum depression more nameable will likely also have downstream effects on other categorization practices, for example, for deciding whether a becoming mother needs medical treatment, etc.

accuracy of our inferential patterns. As the nameability account shows, increasing the nameability of features that are diagnostic of *F* (including *F* itself) might have this effect.[14]

Nameability effects on categorization are also relevant to hermeneutic injustice (Fricker, 2007). First, the connection between nameability and categorization helps *explain* the wrong of hermeneutic injustice. Why does it matter if there are gaps in the linguistic resources that minorities and oppressed groups need to express their experiences? The evidence reviewed in Sect. 4 shows that this creates a problem that goes beyond the (already problematic) situation of having to rely on lengthy circumscriptions. In the experiment with shapes and surface structures, we saw that people were less likely to base their categorization judgments on surface structure if these structures were less nameable. This suggests that, if a category such as sexual harassment is not lexicalized, then, other things being equal, people will also be less likely to categorize a given behavior *as* sexual harrassment (or as other things that sexual harassment is diagnostic of). The wrong of hermeneutic injustice, then, is not only that those who suffer it are at a loss for words when trying to articulate themselves – it is also that they and their contemporaries are less likely to even recognize and appreciate their needs and experiences in non-communicative situations.

Second, the nameability account shows how conceptual engineering can help to partially *remedy* hermeneutic injustice. As we have seen, conceptual engineering can make relevant features more nameable. This, in turn, has positive effects on how *sensitive* our categorization practices are, i.e., how likely it is that we judge that something *e* is *C*, conditional on *e* being *C*. Because of this connection between nameability and the sensitivity of associated categorization practices, conceptual engineering can be a means of increasing the likelihood that things that matter to the lives of oppressed groups and minorities are being recognized by themselves and their contemporaries.[15]

Similar links can be drawn between categorization and many other stated goals of conceptual engineering. Categorization is a fundamental cognitive task with universal application. How we categorize things has downstream consequences for almost everything: how we behave, how we perceive others and ourselves, and how we reason about the world. By demonstrating the impact of conceptual engineering on categorization, the nameability account provides a domain-general and empirically supported explanation and justification for the practice of conceptual engineering.

ACKNOWLEDGEMENTS

We would like to thank Marcel Kertels, Guido Löhr, Christian Nimtz, Jakob Ohlhorst, Andrea Raimondi, and an anonymous reviewer for feedback on earlier versions of this article. We are also grateful to the participants of the conceptual engineering workshops in Warsaw (06/2023) and Uppsala (06/2023), as well as the participants of the EXTRA.12 research colloquium in Bochum (11/2023), for their helpful comments.

Open access funding enabled and organized by Projekt DEAL.

[14] Of course, the accuracy of these inferences depends just as much on their second premises; the empirical and philosophical literature on essentializing and stereotypical inferences suggests that there are systematic problems with our trust in them, too.

[15] This being said, depending on the case and circumstances, inventing a label for a category may also have negative effects. For example, inventing a label for a previously unlabeled oppressed social group may make it easier for members of a dominant group to target them (but see Ritchie, forthcoming for discussion). None of what we have argued here suggests that it is always or even necessarily better for everything to be as nameable as possible. This is why we were careful to conclude, at the end of Sect. 2, that there are *some* categories such that it would have positive worldly consequences if *at least many* people were more accurate at categorizing things according to them.

FUNDING INFORMATION

Gary Lupyan's work on this paper was generously supported by a grant from the National Science Foundation, NSF-PAC Award #2020969.

ORCID

Steffen Koch https://orcid.org/0000-0002-9360-799X
Gary Lupyan https://orcid.org/0000-0001-8441-7433

REFERENCES

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, *64*, 3–26. https://doi.org/10.1006/jecp.1996.2332

Barsalou, L. W. & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 129–163). Cambridge University Press.

Belpaeme, T., & Bleys, J. (2009). The impact of statistical distributions of colours on colour category acquisition. *Journal of Cognitive Science*, *10*(1), 1–20. https://doi.org/10.17791/jcs.2009.10.1.1

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Boutonnet, B. & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *Journal of Neuroscience*, *35*(25), 9329–9335. https://doi.org/10.1523/JNEUROSCI.5111-14.2015

Borghi, A. M. (2020). A future of words: Language and the challenge of abstract concepts. *Journal of Cognition*, *3*(1), 42. https://doi.org/10.5334/joc.134

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. M. (1992). Verbal recoding of visual-stimuli impairs mental image transformations. *Memory & Cognition*, *20*(4), 449–455. https://doi.org/10.3758/BF03210929

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One*, *5*(5), https://doi.org/10.1371/journal.pone.0010773.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Cappelen, H. (2018). *Fixing language. An essay on the foundations of conceptual engineering*. Oxford University Press.

Cappelen, H. (2023). *The Concept of Democracy. An Essay on Conceptual Amelioration and Abandonment*. Oxford University Press.

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

Croft, W., & Cruse, A. D. (2012). *Cognitive Linguistics*. Cambridge University Press.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, *398*, 203–204. https://doi.org/10.1038/18335.

Dembroff, R. (2016). What is sexual orientation? *Philosophers' Imprint*, *16*(3), 1–27. Permalink: http://hdl.handle.net/2027/spo.3521354.0016.003

Dembroff, R. & Wodak, D. (2021). How much gender is too much gender? In J. Khoo & R. K. Sterken (Eds.), *Routledge Handbook of Social and Political Philosophy of Language* (pp. 362–376). Routledge.

Deutsch, M. (2020). Speaker's reference, stipulation, and a dilemma for conceptual engineers. *Philosophical Studies*, 177, 3935–3957. https://doi.org/10.1007/s11098-020-01416-z.

Edmiston, P. & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93–100. https://doi.org/10.1016/j.cognition.2015.06.008

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

Foundalis, H. E. (2006). *Phaeaco: A cognitive architecture inspired by bongard's problems* [PhD. Thesis]. Indiana University.

Fricker, M. (2007). *Epistemic injustice – power and the ethics of knowing*. Oxford University Press.

Gainotti, G. (2014). Old and recent approaches to the problem of non-verbal conceptual disorders in aphasic patients. *Cortex*, *53*, 78–89. https://doi.org/10.1016/j.cortex.2014.01.009
Gennaioli, N. & Shleifer, A. (2010). What comes to mind. *The Quarterly Journal of Economics*, *125*(4), 1399–1433. https://doi.org/10.1162/qjec.2010.125.4.1399
Habgood-Coote, J. (2019). Stop talking about fake news! *Inquiry*, *62*(9–10), 1033–1065. https://doi.org/10.1080/0020174X.2018.1508363
Haslanger, S. (2012). *Resisting reality: Social construction and social critique*. Oxford University Press.
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
Koch, S. (2024). How words matter. A psycholinguistic argument for meaning revision. *Mind & Language 39*(3), 364–380 https://doi.org/10.1111/mila.12494
Koch, S., Löhr, G., & Pinder, M. (2023). Recent work in the theory of conceptual engineering, *Analysis*, *83*(3), 589–603. https://doi.org/10.1093/analys/anad032
Langland-Hassan, P., & Davis, C. P. (2023). A context-sensitive and non-linguistic approach to abstract concepts. *Philosophical Transactions of the Royal Society B.*, *378*(1870). https://doi.org/10.1098/rstb.2021.0355
Leslie, S. J. (2017). The original sin of cognition: Fear prejudice, and generalization, *Journal of Philosophy*, *114*(8), 393–421. https://doi.org/10.5840/jphil2017114828
Lewis, D. (1986). *On the Plurality of Worlds*. Blackwell.
Löhr, G. & Michel, C. (2023). Conceptual engineering, predictive processing, and a new implementation problem. *Mind & Language*, *39*(2), 201–219. https://doi.org/10.1111/mila.12471
Lupyan, G. (2009). Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review*, *16*(4), 711–718. https://doi.org/10.3758/PBR.16.4.711
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., & Mirman, D. (2013). Linking language and categorization: Evidence from aphasia. *Cortex*, *49*(5), 1187–1194. https://doi.org/10.1016/j.cortex.2012.06.006
Lupyan, G., & Zettersten, M. (2021). Does vocabulary help structure the mind? *Minnesota Symposia on Child Psychology*, 160–199. https://doi.org/10.1002/9781119684527.ch6
Lupyan, G. & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170–186. https://doi.org/10.1037/a0024904
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.
Machery, E. (2009). *Doing without concepts*. Oxford University Press.
Machery, E. (2017). *Philosophy within its proper bounds*. Oxford University Press.
Manne, K. (2017). *Down girl. The logic of misogyny*. Oxford University Press.
Munroe, R. P. (2010). *Color survey results*. https://blog.xkcd.com/2010/05/03/color-survey-results/
Murphy, G. (2002). *The big book of concepts*. MIT Press.
Murphy, G. (2023). Categories and concepts. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology*. DEF Publishers.
Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? *Cognition*, *80*(3), 11–20. https://doi.org/10.1016/S0010-0277(01)00112-3
Nimtz, C. (2021). Engineering concepts by engineering social norms: Solving the implementation challenge. *Inquiry*, *67*(6), 1716–1743. https://doi.org/10.1080/0020174X.2021.1956368
Perry, L. K., & Samuelson, L. K. (2013). The role of verbal labels in attention to dimensional similarity. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuch (Eds.), *Proceedings of the thirty-fifth annual conference of the cognitive science society* (pp. 3217–3222). Cognitive Science Society.
Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665–81. https://doi.org/10.1016/j.cognition.2007.04.003
Riggs, J. (2019). Conceptual engineers shouldn't worry about semantic externalism. *Inquiry*, online-first, 1–22. https://doi.org/10.1080/0020174X.2019.1675534
Rissman, L., Liu, Q., & Lupyan, G. (2023). Gaps in the lexicon restrict communication. *Open Mind*, *7*, 412–434. https://doi.org/10.1162/opmi_a_00089

Ritchie, K. (2021). Essentializing language and the prospects for ameliorative projects. *Ethics 131*(3), 460–488. https://doi.org/10.1086/712576

Ritchie, K. (forthcoming). Labeling unlabeled identities. In M. G. Isaac, S. Koch, & K. Scharp (Eds.), *New perspectives on conceptual engineering vol. 3: Applied conceptual engineering*. Springer.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*(3), 369–398. https://doi.org/10.1037/0096-3445.129.3.369.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.). *Cognition and categorization*. Hillsdale: Erlbaum. Reprinted in: E. Margolis and S. Laurence (Eds.) (1999). *Concepts: Core readings*. MIT Press.

Schnell, L. (2020). Coronavirus and social distancing: Why people won't avoid each other. USA *Today*, March 17.

Schwanenflugel, P.J. (1992). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.). *The psychology of word meanings* (pp. 223–250). Lawrence Erlbaum.

Sterken, R. (2020). Linguistic intervention and transformative communicative disruptions. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual Engineering and Conceptual Ethics* (pp. 417–434). Oxford University Press.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1140–1154. https://doi.org/10.1037/0278-7393.21.5.1140

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315. https://doi.org/10.1037/0033-295X.90.4.293

Wodak, D., Leslie, S.-J., & Rhodes, M. (2015). What a loaded generalization: Generics and social cognition. *Philosophy Compass*, *10*(9), 625–635. https://doi.org/10.1111/phc3.12250

Yendrikhovskij, S. N. (2001). Computing color categories from statistics of natural images. *Journal of Imaging Science and Technology*, *45*(5), 409–417.

Zettersten, M. & Lupyan, G. (2020). Finding categories through words: More nameable features improve category learning. *Cognition*, *196*, 104135. https://doi.org/10.1016/j.cognition.2019.104135.

Zettersten, M., Bredemann, C., Kaul, M., Ellis, K., Vlach, H. A., Kirkorian, H., & Lupyan, G. (2023). Nameability supports rule-based category learning in children and adults. *Child Development*, *95*(2), 497–514. https://doi.org/10.1111/cdev.14008

How to cite this article: Koch, S., & Lupyan, G. (2024). What is conceptual engineering good for? The argument from nameability. *Philosophy and Phenomenological Research*, 1–18. https://doi.org/10.1111/phpr.13110