COPiL CAMBRIDGE OCCASIONAL PAPERS *in* LINGUISTICS
Volume **15**, Article 5: pp. 155–195, 2023 | ISSN: 2050-5949

Nullius In Verba
Conceptual Engineering and the Royal Society Corpus*

Nina Haket
University of Cambridge

Abstract Conceptual Engineering is the practice of improving the concepts we use for a specific purpose. However, despite involving words and their meanings, this practice has not been looked at from the perspective of linguistics. This paper takes a small, niche scientific community, namely the Royal Society, and investigates to what extent Newton's proposed distinction between *mass* and *weight*, which can be thought of as an instance of Conceptual Engineering, was consistently used in scientific journal articles between 1700 and 1920. If successful, the scientific senses proposed by Newton should differentiate *mass* and *weight* rather than any other extra-linguistic factors. 1500 tokens of *mass* and *weight* from the Royal Society Corpus were annotated for their sense along with other linguistic and extra-linguistic features, and this data was consequently analysed using Ctrees and Random Forests. The results showed that sense was not an effective predictor of the lemma, but that extra-linguistic factors such as the author and academic field had greater predictive power. It was concluded that Newton's Conceptual Engineering of *mass* and *weight* was unsuccessful. This work shows future research on more diverse communities with less coherent language ideologies is necessary when investigating cases of Conceptual Engineering, and also that linguists can be important contributors to Conceptual Engineering research in the future.

1 Introduction

Conceptual Engineering (CE) is becoming a veritable buzzword within studies on the philosophy of language. CE is the practice of improving the concepts we use for a specific purpose. Rather than taking a concept and analysing what it is and how it is used, as is the case in conceptual analysis, conceptual engineers take a concept, assess it to find defects, and offer an improved concept that addresses these defects, consequently making it more suitable for purpose. CE has recently become hugely popular within philosophy but concerns about concept suitability can already be found in the work of Nietzsche, who suggests 'What is needed above all is an absolute scepticism toward all inherited concepts' (Nietzsche 1901: 220-1). Despite its long history and prominence in philosophy, few linguists are aware of CE.

* Thanks to Lauren Fonteyn and Janet Connor for their wonderful guidance working on the thesis that this article is based on, and to Marina Terkourafi for incredibly helpful feedback upon submission. I am also grateful to Kasia Jaszczolt for further commentary and guidance. Any remaining errors are my own.

Upon being told that CE is the practice of improving our concepts, linguists may be perturbed by the subject of discussion. As linguists, we are descriptive rather than prescriptive. We investigate how language is used rather than how it should be used. This is intrinsically in opposition to CE, which aims to provide prescriptive and normative judgements on how we should be speaking. On the surface, this suggests the two are incompatible.

A recent expansion of CE from simply a philosophical method to a tool for potential societal, moral, or political improvement has changed the success criteria. It no longer suffices to simply suggest an improved concept. Instead, the aim is society-wide adoption and reform. The practice of CE has, therefore, introduced an implementation problem (Cappelen 2018, Deutsch 2020, Jorem 2021, Koch 2021). Any re-engineered or ameliorated concepts must spread through society, but little work has been done on how to spread them, or what patterns we may expect to see.

However, this might be because the question lies outside the domain of philosophical thought and crosses into the domain of linguistics. This work aims to take the first step in analysing the spread of conceptually engineered terms with two intentions. The first is to apply linguistic tools to the issue of concept diffusion, and the second is to draw the attention of linguists and hopefully inspire more work from linguists in CE.

Rather than tackling a large, societal instance of CE, many of which are still ongoing and hard to evaluate, this study looks at CE in the diachronic scientific domain where ameliorating the concept aimed to improve scientific precision. Furthermore, the community in question has clear criteria for whether a concept is defective and definitive authority on who decides what constitutes an improvement. More specifically, I target the usage patterns of *mass* and *weight*, as first differentiated by Newton (1687), between 1700 and 1920 using data from the Royal Society Corpus (RSC) (Fischer, Knappen, Menzel & Teich 2020). Influences on lemma choice may include the author, scientific sub-discipline, and various linguistic factors such as sense, plurality, and part-of-speech. This study aims to use these factors to create a first picture of how engineered terms diffuse, including how long it takes for adoption to occur and what linguistic and social factors can encourage or hinder uptake. These patterns may help evaluate the likelihood of success of CE projects. The specific research questions addressed by this work are the following:

i. To what extent are *mass* and *weight* successfully differentiated by sense after Newton proposed separating the concepts in 1687?

ii. What time frames and trends are there in the diffusion of *mass* and *weight*?

iii. What effect do extra-linguistic variables such as author and subfield have on the choice of lemma?

1.1 Outline

This paper will be structured as follows. Moving forward, section 2 introduces Conceptual Engineering, and section 3 presents the case study used for the investigation,

while section 4 summarises the ideological and theoretical background discussed. The methodology and hypotheses are detailed in section 5 and section 6, and section 7 presents the results. Finally, section 8 discusses the results, and section 9 concludes.

2 Conceptual Engineering

The definition of CE is hard to pin down. As a form of introduction to the topic, the characterisation of Eklund (2014) is a good place to start:

> *[W]hile philosophers often have been concerned with our actual concepts or the properties or relations they stand for, philosophers should also be asking themselves whether these are the best tools for understanding the relevant aspects of reality, and in many cases consider what preferable replacements might be.* (Eklund 2014: 293)

Eklund's statement reveals two things about CE, the first being that it involves finding 'preferable' replacements, also called improvements or ameliorations, for our concepts. Secondly, it shows that CE contrasts with a purely descriptive approach, which aims to find what our concepts are and how we use them. An 'ameliorative' approach, on the other hand, looks to improve these concepts based on any defects we find. Amelioration in CE contrasts with the linguistic definition of amelioration, which refers to a semantic change in which a word gains a more positive meaning. For example, the adjective *nice* previously meant FOOLISH, whereas now it means PLEASANT (Altakhaineh 2018, "nice n.1" 2022: Oxford English Dictionary). In CE, amelioration refers only to improving a concept for a specific purpose.

The assumptions of CE are, therefore, that the concepts we use can be defective, and if found to be defective, we should create alternatives to fix them (a sentiment echoed in the title of Cappelen 2018, *Fixing Language*). However, what is not clear from these definitions is who decides what constitutes a defect and what constitutes an amelioration within CE. For whom are certain concepts preferable, and who has the power to determine that we should change the way we speak? Answering these questions requires us to appeal to language ideologies (Silverstein 1979) and the notion of authority over language. What I aim to show by mentioning this here is that CE is controversial, nor is it clear what is considered an improvement or by whom. It is not a clearly defined and accepted process. This work does not intend to defend nor promote CE as a process but simply aims to show possible contributions from linguistics.

However, if we accept that someone, or a certain body, has the power to determine what counts as a defect in language, the next question is what these defects can be, and there is no singular answer to this. Furthermore, we must consider that amelioration can be relative to a goal, as the function of a concept also needs to be highlighted.

One approach discussed in the literature as being one of the first instances of CE is Carnap's explication (Carnap 1947, 1950, 1962). Carnap was concerned with the

usage of language within science, both in the sense of creating artificial languages and exploring notions of precise language. The latter idea originated from the fact that he believed ordinary concepts to be too vague for precise scientific inquiry (Cull 2021). Put more simply, Carnap proposes replacing vague and imprecise ordinary concepts with more precise ones that fit into a 'broader scientific picture' (Cull 2020: 13), with the criteria that they must be fruitful, simple and exact, as well as being similar to the original term.[1] Carnap, therefore, is engaged in a form of CE, with the idea that concepts can be defective for scientific enquiry.

Since Carnap, it has become accepted that non-epistemic values warrant consideration when engineering concepts, such as the needs and wants of the community, thereby introducing moral, political, and social influences. The inputs to CE thus become plural in the addition of this new dimension. Cappelen & Plunkett (2020) offer a selection of conceptual defects mentioned in the literature, including cognitive defects, moral/political defects, epistemological defects, and metaphysical defects. Once these defects have been identified, conceptual engineers can either replace, revise (potentially by splitting the existing concept into multiple different concepts) or eliminate the defective concept. Furthermore, philosophers can choose to either keep or change the lexical item associated with a word.

For example Haslanger (2000) and Clark & Chalmers (1998) all maintain the original lexical items *woman* and *belief* but choose to improve the concepts WOMAN and BELIEF behind it, whereas Scharp (2013) and Nado (2021) propose multiple concepts to replace the single original 'jack of all trades concepts' (Nado 2021: 2) of TRUTH and KNOWLEDGE. Habgood-Coote (2020) represents the last option, which is to admit the concept is beyond saving and propose getting rid of it altogether. These represent the three strategies mentioned in Cappelen (2018: 35).

A further difference between explication and these newer projects is that what many conceptual engineers are interested in changing is not only the concept but phenomena in the real world (Nado 2021). However, the relationship between language, thought, and the real world is far from simple. CE projects sometimes assume that by changing our concepts we can change either how we perceive the real world or change the world itself. This is an important area that has been explored by many conceptual engineers, and philosophers alike (Koch & Ohlhorst 2022). The implications and theory behind the idea of changing the world through concepts fall beyond the scope of this paper; for now, it suffices to say that even though this is the assumption of many CE projects, it is controversial.

2.1 A note on concepts

Across the literature, it seems as if every article takes a different stance on what concepts are in CE. For the sake of space constraints, a full evaluation is not possible. Instead, this work will take the view of Cull (2021). He suggests that it is possible to sidestep any discussion of concepts if we take CE to be the 'improvement of our representational devices (whatever those are)' (Cull 2021: 232). In his view,

[1] These desiderata are not uncontested or simple; I point the reader to Dutilh Novaes & Reck (2017) for a deeper discussion.

it is simply not relevant to the sort of project he pursues. This work will remain neutral in this debate and looks at word meanings without making any theoretical claims as to the structure of the underlying concepts. This topic deserves thorough investigation, and this falls beyond the scope of this paper.

The relationship between concept diffusion and word diffusion warrants discussion: the two are not the same. As mentioned, the relationship between words and concepts is contested, and so a spread of a certain word with a certain meaning does not mean the spread of a certain concept. However, in examining a historical case, concepts are not accessible, whereas words are. I acknowledge this is not the same, and hope to be able to investigate the relationship between the two in the future.

2.2 The implementation problem

While the issues above are important and fascinating, they are currently (and rigorously) being debated. Now that the reader is broadly aware of these issues, it is possible to turn to the issue that will form the frame of this paper.

Authors who are pessimistic about CE often express doubts due to the impossibility of implementation. CE, therefore, has an implementation problem. Numerous philosophers discuss this problem, with both positive and pessimistic conclusions. Much of the rest of the literature focuses either on potential strategies in the implementation process (Sterken 2020, Andow 2021, Jorem 2021) or on whether implementation is even possible (Koch 2021, Cappelen 2018, Richard 2020, Deutsch 2020).

The problem with all these analyses (except for perhaps Andow 2021) is that they stay in the theoretical domain; they look at potential ways of implementing conceptual changes and debate what is meant by implementation but do not look at concrete examples. As stated by Jorem, 'Neither Cappelen (2018), (Koch 2018) nor Deutsch (2020) assesses particular languages for the feasibility of effecting semantic change in them' (Jorem 2021: 192).

Assessing a concrete instance of CE is a necessary contribution to these debates. While many of the discussions concern whether CE can be implemented, or some vague strategies to do so, many more aspects are important. For example, how long implementation takes, what factors govern it, who adopts the term early, forces of resistance and preferable contexts. Such a study is beyond the scope of philosophy but within the grasp of linguists.

2.3 Linguistics and Conceptual Engineering

What has been discussed up to now is largely prescriptive. CE is about what concepts should be and how we can improve them, something that is strongly against the descriptive principles of linguistic norms. Linguists describe how language is used, how it is structured, and what words mean. It, therefore, seems hard to find anywhere in CE that linguists could contribute. However, descriptive approaches to CE can help find out what our existing concepts are, establish the relevant normative constraints, and establish what kind of concept would best meet these constraints for the relevant

group (Andow 2020: 14). More specifically, linguists can use the descriptive and investigative frameworks they have built to address the implementation problem.

In this solution, linguists need not offer any normative judgements. Language change and how it spreads are major themes in linguistics (Labov 2001, 2010). This work takes a first step in encouraging linguists to participate in CE. The investigation here will target which groups and individuals use the new concepts, how fast and complete the spread is, and in which contexts they are favoured. Of course, this is not everything that linguists can do, but it is a start.

3 Case study

Of all the places a linguist could start with investigating CE, *mass* and *weight*, or scientific literature and concepts more generally, between 1700-1920 do not form an intuitive starting point. However, it provides an ideological background suited to a CE study. The following sections will look in detail at why.

3.1 Newton's England

Around the time Newton was writing, language came to the forefront of public concern (Woolard & Schieffelin 1994). For example, in the *Novum Organum* (Bacon 1623), Bacon (1561-1626) writes about the idol of the marketplace, which consists of words that 'lead men away in numberless empty controversies and idle fancies' (Bacon 2011: 55).[2] In fact, according to Bacon, language in its current state was the greatest hindrance to scientific progress (Bauman & Briggs 2003). Locke possessed a similar viewpoint, referring in his work *An Essay Concerning Human Understanding* (Locke 1689) to the 'abuse of words' (Locke 1975: 508), and the notion that people 'make them stand sometimes for one thing, and sometimes for another' is 'great folly, or greater dishonesty' (Locke 1975: 495-2). In his view, the words one used had the power to shape thought, and therefore faulty words could lead to faulty thoughts (Losonsky 2021).

Ruíz (1984) suggested groups' ideologies could view language as either a resource, problem, or right. Both Locke and Bacon seem to fall within the category of seeing language as a problem ('the cheat and abuse of words') or something that is hindering scientific and human progress. CE generally appears to fit into the problem ideology and therefore aligns with the viewpoints on language throughout time. It seems logical that an ideology of language as a problem, as well as an ideology that we have the power to fix it, is a driving force behind CE.

3.1.1 The Royal Society

The Royal Society started as an 'invisible college' of scientists in 1660 (McDougall-Water, Moxham & Fyfe 2021). Soon after, it gained royal approval and became *The Royal Society of London for Improving Knowledge*. This society allowed fellowships

[2] While the general citations for this work are the original 1620 text, any quotes are taken from more recent translations or edited works, and these will be cited when mentioning the relevant quotations. The same is true of any citations pertaining to Locke below.

for scientists and had prominent members, including Einstein, Newton, and Darwin. The first issue of the journal *The Philosophical Transactions of the Royal Society* was published in 1665 by Henry Oldenburg, the first secretary, and is now the oldest continuously published scientific journal in the world (McDougall-Water et al. 2021). This journal was published monthly and contained scientific correspondence, reviews, experiments, observations, and discussions. Alongside this journal, the society also published other scientific works, including Hooke's *Micrographia*, Newton's *Principia*, and Franklin's kite experiment. In terms of language, The Royal Society turned against the contemporary gentlemanly rhetoric and manipulation of meaning (Dawson 2007), instead rejecting 'all the amplifications, digressions, and swellings of style' (Sprat 1667: 111-13).

The motto of the Royal Society was *Nullius in Verba* ('take nobody's word for it'), encouraging members not to accept scientific discoveries at face value, and promoting the distrust in language and our words. It is for these reasons that the title of this work includes the motto of the Royal Society.

3.2 Newton's mass *and* weight

Within this environment, Newton (1643-1727) proposed the distinction between *mass* and *weight*. Originally, the word *weight* meant 'portion or quantity weighing a definite amount. Often preceded by an expression indicating the amount' ("weight n.1" 2022: OED). The word *mass* does predate Newton's engineering; the meaning 'dense aggregation of objects having the appearance of a single, continuous body' is traced back to 1382 ("mass n.1" 2022: OED) and 'a coherent body of matter of unspecified or indeterminate shape, and usually of relatively large bulk; a solid and distinct object occupying space' can be traced back to 1425. Newton did not create the word.

However, as science progressed, it became important to differentiate the quantity or portion of matter of an object and its downward force due to gravitation. It became necessary to differentiate the property (mass) and the force (weight) acting upon the property. Differentiating the two is needed to measure acceleration using the famous $F = m\,a$ equation (force equals *mass* times acceleration), Newton's Law of Universal Gravitation (Newton 1687), and engineering concepts such as *breaking weight* and *weight-specific resistance* just to name a few.

However, both of these concepts were lexicalised as *weight*. The word *mass* was in existence, but not with its modern scientific meaning; rather it had a collective meaning, or simply referred to an object.

In a first unfinished draft of the *Philosophiæ Naturalis Principia Mathematica* (Newton 1687), henceforth *Principia*, Newton defined *weight* as 'the quantity or amount of matter being moved, apart from considerations of gravity, so long as there is no question of gravitating bodies' (Newton 1999: 87). Newton, therefore, is making explicit that while he is using the term *weight*, he is looking for a way to refer to *mass*. He just simply does not have a lexicalisation for the concept yet. He is searching for a way to discuss the invariable property of the quantity of matter that makes up an object, and not the variable property of *weight*. He is unhappy with

the use of *weight* for this concept and knows that it is unsuitable for discussions that do not involve gravity or 'heaviness' but has to make do due to 'the want of a suitable word' (Newton 1999: 87).

By the first finished draft, around 1685, Newton has settled on a lexicalisation for this concept, namely *mass*. Definition 1 in the *Principia* states that *mass*, and also what he coins *body*, is 'the quantity of matter' that stems from 'its density and volume jointly'. After its first appearance in Newton's work, the first attestation in an English test is in the work of Harris in 1704.

(1) *Hanc autem quantitatem sub nomine corporis vel massæ in sequentibus passim intelligo.*

"I mean this quantity whenever I use the term 'body' or 'mass' in the following pages" (Newton 1687: 15)

(2) *Masse, this Word is used by the Natural Philosophers to express the Quantity of Matter in any Body.* (Harris 1704)

To summarise, *mass* is the amount of matter that makes up an object. *Weight* is the force of gravity acting on the *mass*. *Weight* is proportional to *mass* and acceleration, where acceleration is the acceleration of a free-falling object in that gravitational field. Objects always have *mass* but do not always have weight. For example, an object in free fall is weightless but still has *mass*.

Again, we must stress that Newton did not create the word *mass*, but rather re-allocated concepts among existing lexicalisations. Initially, the word *weight* was used for both concepts, but post-Newton, they achieved their modern scientific meanings.[3]

3.3 Conceptual Engineering?

Mass in its scientific sense is, therefore, a relatively new concept, especially compared to *weight*. *Weight* is a word with much more history and much more varied usage. However, the splitting of *weight* into *mass* and *weight* does not necessarily mean it is an instance of CE. This work, however, takes the opinion that it is, for the reasons specified below.

3.3.1 Carnapian explication

Carnapian explication targets scientific concepts to make them more suitable for purpose. While the everyday vernacular does not need differentiation between *mass* and *weight*, it becomes crucial in the scientific domain.

To recall explication, Carnap writes 'By the procedure of explication, we mean the transformation of an inexact, pre-scientific concept, the *explicandum*, into a

[3] Both terms have undergone shifts due to more recent scientific discoveries, for example, different kinds of mass (Jammer 1997).

new exact concept, the *explicatum*' (Carnap 1947: 3). In this case, *weight* is the pre-scientific lemma, used for both concepts until 1687. The similarity criterion is fulfilled since *mass* and *weight* can be used in the same contexts as *weight* used to be used. The separation into these two concepts rather than a single one allows for more exact scientific enquiry, as shown by Newton's laws and discoveries after he explicitly separated them. This also lends credence to the idea that separating these concepts leads to increased fruitfulness. Simplicity is a difficult criterion, but it could be argued that *mass* and *weight* allow for increased simplicity since they separate matter and the forces acting on it, rather than considering both at once. Newton's differentiation of *mass* and *weight* can therefore be considered as an instance of Carnapian explication, one of the most frequently discussed forms of CE.

3.3.2 Fixing a problem

CE is centred around the fact that concepts can be used to fix a problem. As mentioned, Newton found the ambiguity of *weight* to be problematic, and therefore offered a separation into *weight* and *mass* as an amelioration. He clearly identified a problematic concept, or the fact that one lexical item contained multiple concepts, and proposed a solution to fix it. This appears to fit nicely into the study of CE.

4 Interim summary

As discussed, language ideologies and authority provide a complicated barrier to understanding what is considered an amelioration and who has the power to propose ameliorated concepts.

The scientific community has one overarching goal: scientific progress. The scientific register has a definitive authority, namely, those who can prove their theories, and the viewpoints of many scientists within this sphere are well-documented. An ameliorated concept is thus one that allows for scientific progress, potentially by being more precise or accurate.

Authority was held by those in charge of the society, and also those who conducted respectable scientific research, thereby creating a somewhat Habermasian society in that it responded to 'the power of the better argument'. The Royal Society participated in the creation of the scientific register, which stood in contrast with the vernacular of the time. Philosophical approaches towards language at the time stressed the unreliability of language, especially regarding its precision in scientific study.

The engineering of *mass* and *weight* within this framework fits with the ideological conceptions of amelioration and defects within the Royal Society. The scientific register provides the data used to identify whether Newton's suggestion to differentiate *mass* and *weight* by sense was successful, or whether other extra-linguistic factors, such as the author or the subfield, had a greater effect.

5 Methodology

Now that we have established that *mass* and *weight* are an instance of a CE project, and covered the ideological background upon which it can be projected, it is possible to detail the methodology of this study.

To operationalise the research questions from section 1, the method chosen needs to identify the factors that condition the choice of either *mass* or *weight* and rank these factors based on their significance. Through doing this, it is possible to identify whether sense conditions the usage of *mass* and *weight*, or whether extra-linguistic factors have a larger influence. In the former case, one could argue for successful CE, and in the latter case, idiosyncratic usage continues to overrule any CE proposals made.

5.1 Concepts, words, and written texts

Concepts exist in the mind, not in language, making the direct, quantitative study of concepts difficult without access to human cognition (Meyns 2020). With historical data, only the linguistic manifestation of these concepts remains, which may not be reliable. This study looks at the linguistic uses of *mass* and *weight* to try and investigate how and when the concepts became separate. Rather than doing this through a study of cognition, this will occur through a study of usage. There is the assumption that if usage patterns start to diverge between these two terms, there is evidence of the acceptance of the differential lexicalisations of these concepts.

Turning now to the type of data used, it is commonly thought that the vernacular is the primary and best object of linguistic research, and that written sources are secondary to this spoken data (Gray & Biber 2018). However, registers differ in important ways that are largely contextually determined. Differing communicative goals can lead to different registers, and the scientific community created aimed to share information among its members, and during this time the main medium for doing so was through written language, especially journal articles and letters. The public associated with the Royal Society was a reading public as opposed to a speaking public. Therefore, in the case of the scientific register from 1700 to 1920, written language is not the next best thing, it is the best option.

5.2 Sub-corpus creation

This study uses the Royal Society Corpus 6.0 (Fischer et al. 2020), which is a linguistic corpus based on *The Philosophical Transactions of the Royal Society* and the other journals of the society. The most recent release (6.0) was in 2020 (Fischer et al. 2020), and spans from 1665 to 1996, of which the data from 1665 to 1920 is available via the CQPweb interface (Hardie 2012) from Saarland University.

The corpus was searched online through CQPweb (Hardie 2012, `https://cqpweb.lancs.ac.uk/`). All instances of *mass*, *masses*, *weight*, and *weights* were identified and downloaded. Extremely low frequencies of some of the non-standard spelling variants were excluded from the data. This gave a total of 56,813 results in the new *weight-mass* sub-corpus.

This collection was then filtered by the frequency with which each author occurred in the sub-corpus to facilitate annotation by hand. Only authors who used *mass* and *weight* over 100 times were included in the present study. This meant that fewer authors were present, and so the chances that a single author contributes more than one token to the dataset when it is randomly sampled is maximised. Authors that occur with a high frequency in the sub-corpus are more likely to use either *mass* or *weight* often in their work, suggesting that these concepts form the theoretical core of their research or that they discuss these terms in some depth. Furthermore, this methodological choice may skew the results since authors studying and focusing on these concepts may be more likely to separate the two concepts than other authors. If even this group is found to have incomplete or inconsistent separation of the concepts, then it is unlikely the lower-frequency authors would have used them consistently.

The sub-corpus consisted of 15,249 total instances after applying the filter.[4] 1500 tokens were taken for annotation. To keep the 1500 tokens representative of the original sub-corpus, the filtered sub-corpus contained 476 random tokens of *mass*, 145 of *masses*, 700 of *weight* and 179 of *weights* to occur with equal frequency compared to the total corpus and to ensure a representative distribution of each author and year. The data, code, and models created by this methodology are available to the reader in Appendix A.3.

5.3 Annotation

The annotation for these 1500 tokens was performed by hand, and included annotations for 'subfield', 'object', and 'sense', in addition to existing metadata from the RSC 6.0 Open.

5.3.1 Subfield

Firstly, tokens were annotated according to 7 'subfield' categories using their titles and the original text: Astronomy, Biology, Chemistry, Engineering, Geology, Physics, and Meta. Meta texts discuss the creation of standard weights and measurements and the weight of ancient coinage and do not fall under any of the other categories posited.

It is worth noting that these are not all mutually exclusive categories, and there is much overlap between them. Texts were annotated by what was perceived to be the main aspect of the contents. The labels given for the 'subfield' are, in a sense, arbitrary and were not designed to be definitive splits between the disciplines. The boundaries we have in place today simply do not apply to historical scientific enquiry, and scientists often worked on subjects that crossed modern boundaries. Furthermore, they provide a macro-topic of the token in question.

[4] *mass* = 4,843 (31.7%), *masses* = 1,475 (9.7%), *weight* = 7,119, (46.7%), *weights* = 1,820 (11.9%).

5.3.2 Sense

The label 'sense' is slightly misleading. Not all of the annotations refer to the sense, or meaning, of the lemma in question. Rather, some of them refer to how the lemmas are used, or the linguistic devices they form part of.[5] All 'sense' annotations are listed in Table 1 below. Instead, it is useful to think of them as general statements on how the lemma is used in context.

The nature of the individual labels is discussed in their respective subheadings below. I will refer to these as 'senses' from this point onward, fully acknowledging that this is not the most accurate label.

To avoid confusion between the lemmas and the senses, the lemmas will be referred to by italics, such as *mass* and *weight*, whereas the senses will be referred to by small capitals, such as MASS and WEIGHT, or the letter M or W. The word *mass* can therefore be used with both the senses MASS and WEIGHT.

Sense	Label	
Mass	M	Used with the scientific sense mass
Weight	W	Used with the scientific sense weight
Nominal	N	Used as a noun, or nominal item in the sentence
Metaphor	MET	Used in a metaphorical context
Calque	Q	Calque or translation of the French phrase en masse
Collective	COL	Used to refer to a collection of objects
Unsure	W/M	Used when unclear if sense is mass or weight

Table 1 Sense annotations.

5.3.3 M

The annotation M refers to the usage of a lemma with the meaning, or concept, of scientific MASS. To revisit the previous discussion, this refers to how much matter is within an object and is usually measured using the (kilo)gram. Both the words *mass* and *weight* can be used with the 'sense' M, as is highlighted below.

(3) *We are thus led to inquire how the stresses are distributed in the earth's mass and what are their magnitudes* (Darwin 1882)

(4) *In the third, the weight of the principle bones of a selected number of species (27) is stated* (Davy 1865)

[5] In hindsight, I would have proceeded differently with the 'sense' annotations. In this study, the 'sense' categories emerged in a bottom-up fashion as I annotated the data. If I were to do it again, I would work in a top-down fashion to ensure coherence in the labels.

In (3), Darwin is talking about the amount of matter within the earth. In contrast,(4) uses the word *weight* but with the 'sense' M, referring to the amount of matter within and between bird species rather than the downward force they generate.

5.3.4 W

W refers to the usage of a lemma with the meaning, or concept, of scientific WEIGHT. As mentioned in the previous section, this refers to the interaction of MASS with a specific gravitational field; it is a measure of force rather than a measure of matter. As such, tokens with the 'sense' W explicitly refer to force, balancing, counterpoises, or the amount of effort required to lift something. While the current official unit of WEIGHT is Newtons, this convention was only introduced in 1948, after the end of the corpus. Before this time, the same units were used for both *mass* and *weight*, meaning this had no contribution to the annotations.

(5) *fig. 3 is only 40 feet from the bow, and that the excess of weight over buoyancy on this length is only 45 tons* (Reed & Stokes 1871)

There are no instances of *mass* having 'sense' W in the annotated corpus. This interesting one-sided polysemy of *weight* but not *mass* will be addressed in the discussion section. (5) shows the word *weight* with 'sense' W. Reed & Stokes are referring to the balance between the upwards force of buoyancy within the water and the downward force generated by the interaction of a ship's *mass* with the gravitational field.

5.3.5 W/M

This annotation is used when it is unclear whether the token refers to the 'sense' M or the 'sense' W. Again, there are no examples of this 'sense' for *mass*; it appears as if it is always clear which 'sense' is being used, which is almost always N or M. It does, however, occur with *weight*.

(6) *The Commissioners for the Restoration of the Standards of weight and measure, in their Report dated December 21, 1841, recommended that…* (Miller 1857)

(6) is an instance of what has previously been described as a Meta text. Miller is writing about the construction of the new imperial standard pound and refers to the commissioners of 'standard weight and measure'. In this case, *weight* is not used as a noun interchangeable with *object* but does refer to a property that objects have. However, they may be talking about standardising the unit of MASS or standardising the unit of WEIGHT. It is not made clear which they speak of. Therefore, the category reflects this ambiguity.

5.3.6 N

The annotation N refers to concrete entities and is not referring to the scientific property of MASS or WEIGHT. Of course, these entities have MASS or WEIGHT, but this property is not what is being referred to, rather the focus is the object itself. As per the discussion in section 5.3.2, this is not a meaning or 'sense' per se. Rather, it is a nominal usage. One way this was judged was whether the lemma could be replaced by *stuff* or *thing* without losing the intended meaning.

(7) *the mass on the filter was treated with boiling alcohol* (Schunck 1853)

(8) *a flat circular weight nicely turned, and pierced in the direction of its diameter to receive the bar, was slid upon it* (Kater 1819)

(7) shows the word *mass* being used to refer to a chemical compound that had appeared on a filter following a chemistry experiment. *Mass* in this instance a noun replaceable by*stuff*. In (8), the word *weight* is used to refer to an object used in scientific experiments, in a similar way that we speak about *weights* at the gym. A nominal usage can often be identified using adjectives or modifiers; in the case of (8), the weight is described as 'flat circular'. The concept of WEIGHT being the force upon the MASS cannot be either flat or circular, but a thing can be.

Since N is not a sense or meaning, the items annotated as N may have senses M or W, causing overlap in the annotations. It was ensured this was not the case. Rather, these tokens usually corresponded with the meaning 'a coherent body of matter of unspecified or indeterminate shape, and usually of relatively large bulk; a solid and distinct object occupying space. (1425)' ("mass n.2" 2022: OED), or 'portion or quantity weighing a definite amount' ("weight n.1" 2022: OED).

5.3.7 MET

A metaphor is not a sense or a meaning. Rather, it is a linguistic device, and it is once again possible that a lemma that is used metaphorically is also used with the sense M or W. Again, care was taken that this was not the case and that there was no overlap. Tokens annotated as MET usually fall under the definitions of 'a burden (of responsibility, obligation, suffering, years, etc.) (1380)' ("weight n.1" 2022: OED). Interestingly, this may be related to sense W due to metaphor creation processes, such as conceptual mapping (Lakoff & Johnson 1981).

Both the words *weight* and *mass* can be used in a metaphorical sense as well as a descriptive or nominal sense.

(9) *The next thought is that I may have assigned too great a mass to the doubt* (Pratt 1854)

(10) *The contact theory has long had possession of men's minds, is sustained by a great weight of authority* (Faraday 1840)

Both (9) and (10) from Pratt and Faraday respectively refer to the *weight* or *mass* in a metaphorical sense. No literal WEIGHT is being described, but rather the importance or impact of a certain authority, statement, or opinion.

5.3.8 Q

Q is also not technically a sense. Instead, it refers to a calque or translation of the French phrase *en masse* which is also used in English, referring to a single body or group operating collectively. It occurs only twice in the 1500-word sub-corpus, and in hindsight, I would have removed this 'sense' label from the final statistical analysis.

5.3.9 COL

The annotation COL refers to the following entry in the OED: 'dense aggregation of objects having the appearance of a single, continuous body. Also figurative. (1382)' and 'a large amount, number, or quantity of a thing or things, material or immaterial (often with the sense of oppressive or bewildering abundance). Now frequently (colloquial) in plural, sometimes with singular agreement. (1566)'. Thus, it is used to refer to a collection of objects rather than the objects themselves, or the property of either MASS or WEIGHT.

5.3.10 Summary of senses

7 senses have been annotated for usage in the statistical analysis. Two of these, M and W are what will henceforth be referred to as the target senses since these are the senses that Newton targeted with his CE project, and are the main target of the present study. The other senses, N, W/M, COL, Q and MET are henceforth the non-target senses.

5.3.11 Object

The 'object' of the token is what the property of MASS or WEIGHT refers to, more specifically what kind of object has the specified MASS or WEIGHT. In cases where the 'sense' is neither MASS nor WEIGHT, the 'object' still refers to what the token represents, or what kind of 'object' the nominal or metaphorical, or scientific sense has. While this remains vague, some of the examples previously used to illustrate the various senses will again be used to explain what is meant by the 'object'.

(11) *The mass on the filter was treated with boiling alcohol* (Schunck 1853)

The *mass* in this case is a chemical of sorts, whether it be the outcome of a chemical reaction, an atom, or a powder used in chemical experiments. Therefore, this token gets the annotation C, for 'Chemical'. Other 'object' annotations can be seen in Table 2.

Many of the 'object' annotations will be more likely to occur within certain 'subfields', causing nesting. For example, the 'object' GEO in Geology, and S in Astronomy. At first glance, this may make the 'object' annotations seem redundant. However, for many of the 'subfields', the 'objects' discussed are diverse. In Physics, one might use *mass* to refer to matter, a weighted object, a vessel or machine, force,

Abbreviation	Object	Description
ABS	Abstract	Not physical, such as importance and impact
BIO	Biological	Biological items such as cells, cancers, and blood
BO	Body	The human body
C	Chemical	Of a chemical nature, whether it be a chemical compound, or the product of reaction
COINS	Coins	Either contemporary or ancient forms of currency
CONCEPT	Concept	The concept of what *weight* or *mass* are
EQ	Equipment	Equipment in scientific experiments, such as test tubes or pendulums
F	Fluid	Fluid, for example water or alcohol
FOR	Force	Downwards force, for example breaking weight or weight specific resistance
GEN	General	Generalising statements about a group of entities
GEO	Geological	Geological in nature, including glaciers, lava, and volcanoes
MAT	Material	The material something is made up of, for example electricity passing through a sheet of iron
MATTER	Matter	Referring specifically to matter
S	Space	Astrological in nature, namely planets and stars
STAN	Standard	A standard of weight or measure
VM	Machine	Forms of transportation, or mechanical devices. Examples include ships and engines
WO	Weighted Object	Objects that have weight and are used in experiments/observations. When the object is WO, the word can be replaced with *thing*

Table 2 Object categories in annotation.

fluid, equipment, a shape, a concept, and so forth. While some 'objects' may be more likely in certain 'subfields' than others, specific 'subfields' do not guarantee the usage of specific 'objects' token in question, standing in contrast to the 'subfield', which functions as the macro-topic. The question remains as to what outcome can be predicted for this variable. By separating these, we reduce the risk of relying on academic disciplines that were not as separate or modular as today. Instead, we have a more fine-grained idea of whether the topic or object discussed can affect the choice of lemma.

'Object' is related to 'sense' in that 'objects' may become more associated with certain words and senses through frequency effects; if certain combinations appear more frequently, they will become more entrenched in the minds of speakers.

5.4 Data Analysis

Thus, once all the annotation was completed, the metadata that will be used in the analysis is as follows: lemma, 'sense', 'object', 'subfield', 'author', and 'year'.

Firstly, a simple comparison of frequency over time is used, both to track the percentage and absolute frequency of each of the terms and all their senses, and to check the engineered terms for 'correctness of usage'.[6]

To assess which variables influence the choice of lemma, Conditional Inference Trees (Ctrees) and Random Forests (Breiman 2001) were both conducted in R using the party package (Hothorn, Buehlmann, Dudoit, Molinaro & Laan 2006a, Hothorn, Hornik & Zeileis 2006b). The code is based on that provided by Fonteyn & Nini (2020). These analyses work by using the independent variables specified to predict which outcome will occur in a specific dataset, in this case, what predictive power each variable has over the choice of lemma. The Ctrees create a visual hierarchy of how the data is split to predict the outcome variable. The Random Forests, made up of multiple Ctrees, are used to create Variable Importance Rankings (VIRs). These quantify and rank the importance of the independent variables in the model in predicting the dependent variable. However, since each variable has multiple levels, or multiple annotations, this will not tell us which of the annotations is contributing the most to the importance. For this, it is necessary to look at the Ctrees themselves, since these show which annotations the model is using to predict the lemma.

The Ctrees and Random Forests will be used to create two kinds of models. The first kind of model will only involve cases where the 'sense' annotation is either W or M, the target senses, to investigate the instances where these senses are used, and whether there is a match or mismatch between the 'sense' and lemma used.[7]

The second kind will include all annotated senses, W, M, N, Q, COL and MET, thus adding the non-target senses to the model. While this is not as effective at identifying the factors that influence the choice of the lemma, it is effective at identifying the patterns in the larger lexical network. If it is found that the word *weight* becomes more closely associated with the 'sense' W over time, and the word *mass* becomes more closely associated with M, it is still not known how this affects the other senses of the terms *mass* and *weight*. We may see entrenchment effects,

[6] Generally speaking, a usage-based approach to language takes grammar to be the culmination of one's language experience (Bybee 2006). The frequency with which the lemmas *mass* and *weight* occur therefore do not only give us a synchronic snapshot of how often these words are used at a specific time. Rather, they form the linguistic input for the next generation of scientists and readers. A higher frequency of usage of *weight* with a certain 'sense', or a certain 'object', will therefore contribute to the entrenchment and salience of certain types and tokens, affecting their diachronic development. Therefore, including imbalances in the frequency of usage of *mass* and *weight* is crucial to model the language knowledge of speakers at the time.

[7] These are chosen since these are the choice contexts; authors can choose to use either the word *weight* or the word *mass* in these contexts interchangeably. For example, 'the *weight/mass* of the table is 12kg' is a context in which authors have the choice between the two terms.

such as the increase in metaphorical or nominal contexts. While *mass* and *weight* may have been engineered to reduce ambiguity in one specific area of scientific discourse, the result may be increased ambiguity or polysemy in another place in the lexical network. These models aim to extract these effects.

6 Hypotheses

In using this methodology to operationalise the research questions, we can now concretely state what we expect to find concerning the research questions.

If the Conceptual Engineering of *mass* and *weight* was successful, then the Random Forests and Ctrees will show 'sense' to be the factor with the highest importance when splitting the data. There would also be an increase in the usage of *mass* and *weight* with their 'correct' respective senses MASS and WEIGHT. If unsuccessful, the Random Forests and Ctrees will instead show that extra-linguistic factors such as 'author' or 'subfield' have higher importance than 'sense' when predicting lemma choice. There would also be no consistent usage of *mass* and *weight* with their 'correct' respective senses MASS and WEIGHT over time.

7 Results

This section will use the data and its annotations described in the previous section to investigate the factors that determine the choice of lemma.

7.1 Frequency of senses

Turning first to the frequency of the different senses over time, Figure 1 and Figure 2 show the normalised frequency of each sense per 100,000 words for both *mass* and *weight*.

This normalisation used Equation 1 to find f_{norm} (the normalised frequency per 100,000 words).

$$f_{norm} = \frac{N_{sense} T_{period}}{A_{lemma}} \frac{100,000}{S_{sub-corpus}} \tag{1}$$

Where N_{sense} is the number of tokens of lemma with sense in question in a particular time period, T_{period} is the total number of tokens of lemma in the time period, A_{lemma} is all the tokens of the lemma in the entire corpus in the time period, and $S_{sub-corpus}$ is the total number of words in the time period in the annotated sub-corpus.

First noticeable from Figure 1 is that WEIGHT is not the sense used most frequently with *weight*. If this was a successful instance of CE, it would be expected to see an initial phase of fluctuation of senses before *weight* and *mass* were engineered by Newton. Even for a short while afterwards, fluctuations would be expected. However, in the long run, it would be expected to see a steady increase in the proportion of *weight* used with the sense WEIGHT. This is not the case.

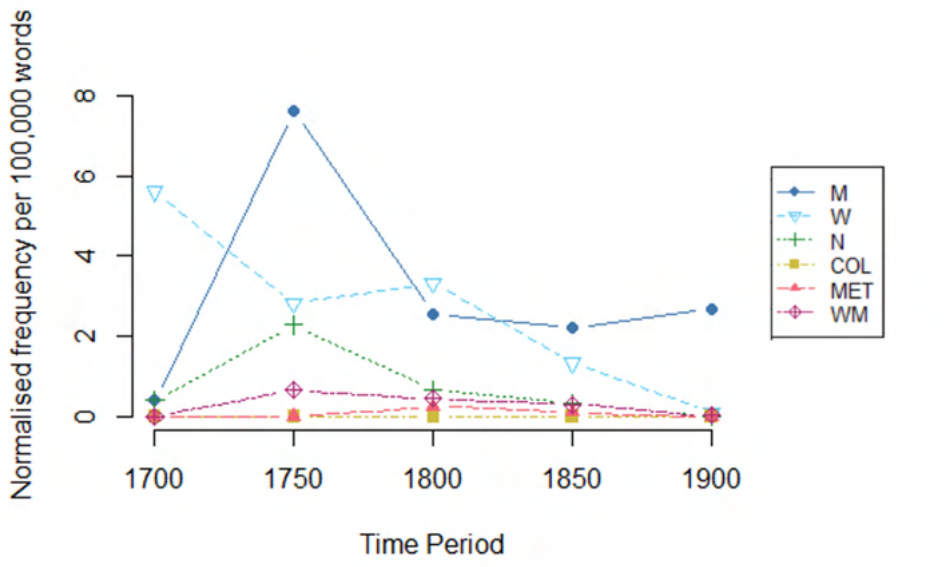

Figure 1 Frequency of all senses of *weight* over time.

The period 1700-1750 looks like an anomaly on the graph in comparison to the trend over time. While MASS is the dominant sense throughout nearly the whole period, in this initial phase it appears with a much lower proportion, and WEIGHT is dominant. However, this is because there are very few tokens in this period in the sub-corpus, and all of them were written by a single persion: Desaguliers It is, therefore, not possible to draw statistically sound conclusions that this was indeed the dominant trend at the time.

However, it does allow for an interesting observation.Desaguliers was an avid Newtonite, even writing a book named *The Newtonian System of the World, the Best Model of Government: An Allegorical Poem* in 1728. The uptake of Newton's engineered terms by Desaguliers could reflect Desaguliers' language ideologies, framing Newton as an authoritative inspiration to scientific progress. The period at the start of the 18th century coincides with the time that Newton was the president of the Royal Society (1703-1727), meaning he would have had large amounts of influence, authority, and control over what got published, especially when it came to the representation of his ideas and theories. Banks (2008) notes that Newton controlled the Royal Society in a highly authoritarian manner.

It is correct to be sceptical that all tokens from this period are from Desaguliers. However, Desaguliers is the only author to appear in the corpus for this period once the filter was applied to find authors who used the words *mass* or *weight* over 100 times. It would be interesting to see if the same patterns emerge in the entirety of the RSC.

1800-1825 is the only time for which the data supports the idea that Newton's engineered terms are being successfully and accurately used. For this period only, *weight* is used most frequently with the sense WEIGHT rather than with MASS.

However, after this period, the proportion of WEIGHT drops again, and the differences in the proportions become pronounced. Thus, if we measure the success of this CE project by the proportion of *weight* used with the sense WEIGHT, this instance of CE did not succeed.[8]

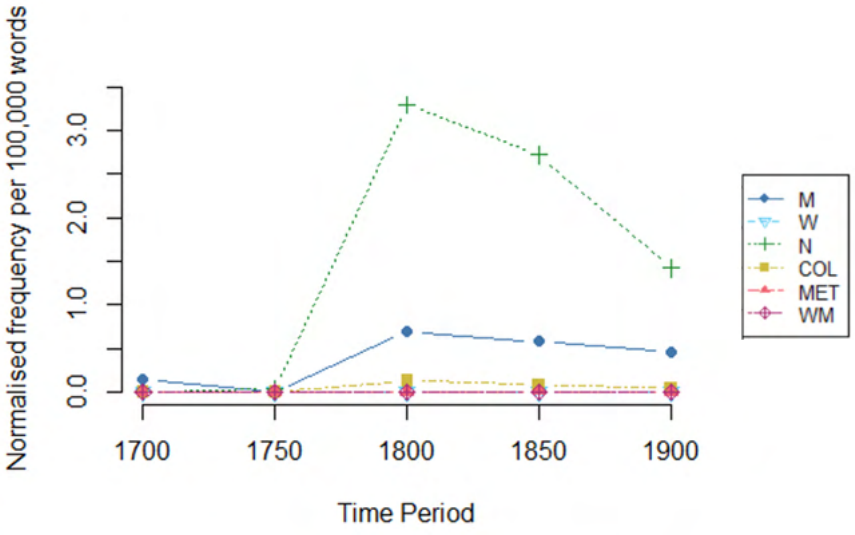

Figure 2 Frequency of all senses of *mass* over time.

Figure 2 shows the same data for *mass*. At the start of the periods in question, the data is sporadic. Despite first being used with the sense MASS in 1704, only after 1750 does the lemma *mass* become frequently used with any of the senses analysed. From this time onward, the nominal sense is by far the most frequent, displaying a slight downward trend but remaining significantly above the other senses. Since this nominal sense of *mass* was established before Newton's CE, this appears to be a continuation of the way it was used before the *Principia*.

Furthermore, while *weight* is used with both senses MASS and WEIGHT, *mass* is never used with the sense WEIGHT. It looks like polysemy in the lemma *weight*, containing both related 'senses' M and W, but homonymy in *mass*, embodying unrelated 'senses' M and N.

7.2 Ctrees and Random Forests

While the frequency data is useful in indicating whether the CE project was successful, another metric we can look at to measure success is the factors that determine the choice between *mass* and *weight*. To get more insight into the diffusion, models were made to check, for each period, which factors help determine the choice between *mass* and *weight*.

[8] It must be said, however, that the usage of the terms was not evenly distributed by author. Due to the historical nature of the study, the authors used different terms in different proportions. Hence, were I to do the study again, it would be necessary to relativise the findings to the authors. After all, a period with only a single author cannot tell us much about whether the term caught on or not.

7.2.1 Models for all data

First, a general picture will be shown of all 1500 tokens over all time periods. The Variable Importance Rankings (VIRs), which rank the variables by their importance in predicting the choice of lemma, can be seen in Figure 3 and Figure 4.

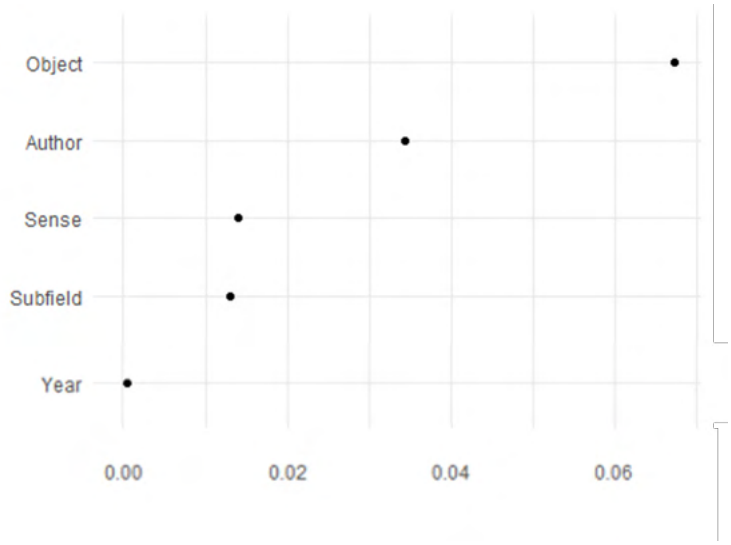

Figure 3 VIR target senses

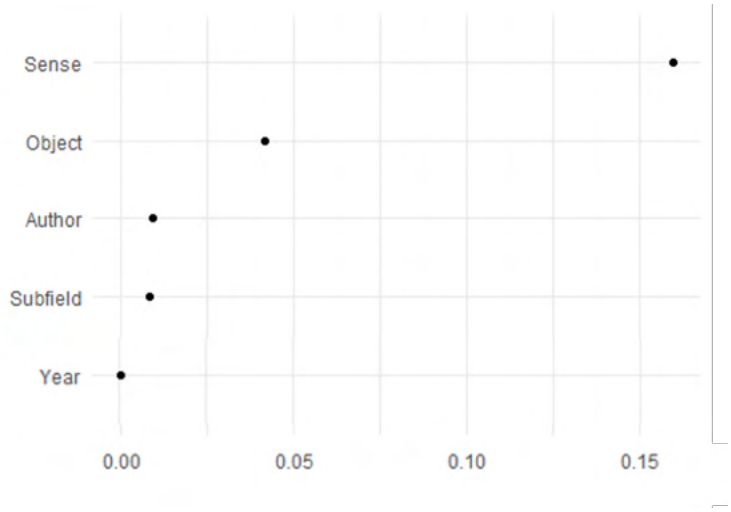

Figure 4 VIR all senses

When only the target senses are considered, as in Figure 3, 'object' ranks most highly, followed by 'author' and then 'sense'. However, when all senses are considered, the VIR in Figure 4 shows that the 'sense' is the most important predictor when all senses are considered, followed by 'object' and 'author'.

This is an interesting discovery since, as is mentioned in the methodology section, there is some nesting when it comes to the 'subfield' and the 'object'. Regardless of this nesting, 'object' seems to be more important than 'subfield' when it comes to predicting whether *mass* or *weight* is used, both for all senses and only the target senses.

'Sense' ranking highly in the VIRs does not tell us which of the annotation labels are contributing most to this importance. However, we can infer that since 'sense' ranks more highly in the VIR for all senses than the VIR for the target senses, this suggests that the senses COL, MET, Q and N contribute to this importance more than M and W, and the addition of the other senses to the model makes 'sense' a much more powerful predictor.

This was confirmed by looking at the Ctree generated on this dataset, which can be found in Appendix A.1. This Ctree showed MASS and WEIGHT, along with MET and W/M, formed a single group that predicted a higher chance of lemma *mass*, whereas COL and N predicted a higher chance of lemma *weight*. This is significant since this is not due to an imbalance in the annotations. N and M are relatively balanced, with 563 and 586 tokens each, with W having 265. Since MASS and WEIGHT are both in the same group and both predict *mass*, it is not the distinction between MASS and WEIGHT that is causing 'sense' to rank highly.

The high ranking of 'author' in both models shows that to predict whether the word *mass* or *weight* is used in the sample, the model has a better chance of making an accurate prediction when it splits the data into different groups of authors than it does when it splits the data according to the 'sense' of the word. What this shows is not that individuals have free choice concerning which word they can use, but that the variation between *mass* and *weight* appears to be individual rather than a community-wide, sense-based phenomenon.

'Object' and 'subfield' provide micro and macro perspectives on the topic of the article and lemma respectively. Both are found to be important in both models, suggesting variation between disciplines and between topics under discussion. Again, variation based on these factors suggests that consistent differentiation based on 'sense' has not been achieved.

'Year', surprisingly, has no predictive power according to this model. The factor 'year' shows whether time helps predict whether *mass* or *weight* is chosen. If *mass* is replacing *weight* in the sense MASS, then the likelihood of *mass* should go up as a function of time. However, the data suggest that such a temporal effect is very weak if it occurs at all. It is possible that 'year' still interacts with *mass*, but that it was not included in this set-up.

In sum, the models created on the entirety of the dataset show that 'sense' is important in the model containing all senses, but that it had lesser importance when it came to the target senses. Other factors that rank highly are 'author', 'object' and 'subfield'.

7.2.2 Time period specific analysis

At first glance, it appeared that time ('year') had little predictive importance. The general model does not, however, indicate whether there may have been any smaller shifts in factor importance across time. To more effectively explore how *mass* and *weight* changed over time, a closer examination of each period is required. This investigation takes intervals of 50 years and looks both quantitatively at the Ctrees and Random Forests created for those years, and qualitatively at any sentences that are of interest.

Figure 5 VIRs for each period.

This analysis was based on the VIRs for each period, for both target senses and all senses, which are shown in Figure 5. These time-specific analyses only cover the periods 1800-1920. In 1700-1750, there is one token of *mass* in contrast to 16 instances of *weight*. There are, therefore, too few tokens to be subject to multifactorial statistical analysis. In 1750-1800, there are only 3 instances of *mass*

opposed to 123 of *weight*. This extreme class imbalance means the model is likely to predict the lemma *weight* in all cases due to the high percentage of accuracy it would deliver. For these reasons, the discussion below will apply only to the periods 1800-1850, 1850-1900, and 1900-1920.

7.2.3 All senses

As can be seen from Figure 5, in the models containing all senses, 'sense' remains by far the most important variable in all three time periods. However, once we have consulted the Ctree to identify which labels are the most important in predicting the lemma, we again see that 'senses' W, M and MET predict usage of the lemma *weight*, whereas N and COL predict *mass*. Full details of the Ctree of the data from between 1850 and 1900 are shown in Appendix A.2. This stands in opposition to what we would want in a CE project; we want the 'senses' W and M to become more differentiated over time, and predict lemmas *weight* and *mass* respectively, and yet this is not seen in the data.

The second and third most important factors in the VIR do fluctuate over time. In 1800-1850 and 1850-1900, they are 'object' and 'author', and in 1900-1920 they are 'subfield' and 'object'.

7.2.4 Target senses

Turning now to the models containing only the target 'senses' W and M, much more variation is found. Between 1800 and 1850, 'author' is the most important, followed by 'subfield' and 'object'. Between 1850 and 1900, 'object' is most important, followed by 'author' and 'subfield', and between 1900 and 1920, 'subfield' is the most important, followed by 'author' and 'object'. Each period, therefore, has a different factor as the most important, but the same three factors are consistently found important in each period.

'Author' always ranks in the top three in each period, which shows that individual preference or choice governs lemma choice more than the 'senses' of the words themselves. The fact that 'author' is ranked more highly than 'sense' suggests that the two-way distinction between *mass* and *weight* is only applicable to certain authors.

The fact that 'subfield' ranks consistently high suggests some differentiation within the different scientific sub-communities. 'Subfield' seems to have gained importance over time, being second in 1800-1850, third in 1850-1900 and first in 1900-1920. This could suggest that the specific scientific subdomains are slowly gaining more power over the choice of lemma.

'Object' is ranked highly across all models, but this may be because of the nesting with 'subfield'. Certain subfields are more likely to discuss certain objects, and so the high ranking of 'subfield' in this model may have affected the ranking of 'object', meaning its ranking must be taken with scepticism. However, in 1850-1900 'object' ranks higher than 'subfield', meaning that this micro-measurement of the topic was more important than the overall sub-discipline in which the lemma is used.

While the 'object' is an important conditioning factor, there does not appear to be any consistency over time concerning which 'objects' particularly condition which lexical item. This goes against the intuition that certain 'objects' may become more associated with certain words and senses through frequency effects. Usage is in such constant fluctuation that consistency is not achieved.

The VIR created on the entire dataset did assign some significance to 'sense' in the model containing the target senses, but this is not reflected here. However, this does not mean that there was no explicit metalinguistic discussion about how the words should be used.[9]

In (12) Crookes is making the statement that he will be using the term *weight* to refer to the 'sense' MASS.

(12) *The weight in grammes is another name for the mass.* (Crookes 1873)

Rather than simply using the term *mass* in this case, Crookes has specified that he will be using one term to refer to the other. Crookes explicitly states what he will be using the term *weight* to convey, leaving none of the ambiguity that the CE project was designed to combat.

The 'sense' MET does not occur with any particular frequency, and so the frequency analyses from section 7.1 do not shed much light on the metaphorical usage of the two lemmas. However, the usage of both these terms in metaphors is a case of growing polysemy that is theoretically very interesting when it comes to investigating CE. (13) shows the sole instance of *mass* being used with a metaphorical sense.

(13) *I may have assigned too great a mass to the Doubt* Some idea may be formed of the amount of these deflections* (Pratt 1854)

This kind of construction would occur more frequently with *weight*, such as 'What I more especially lay weight upon is this' (Kopp 1865). More instances use *weight* in this construction, all between 1822 and 1894, and this irregular metaphorical usage of *mass* can be found right in the middle of this period of high usage of *weight* with this sense. This may potentially be an instance of a kind of transfer; since *weight* was being used in this way frequently, and there were many contexts where *mass* and *weight* are used interchangeably, the concepts may have started to be perceived as linked (or at least by this author). Changes affecting *weight*, namely the increase of metaphorical usage, are spread to *mass*, potentially in a network-like fashion. This is an interesting area of future research to be done with the entirety of the RSC.

7.3 Summary

Returning to the hypotheses, we can see that 'sense' is not the most important variable when it comes to the models containing the target senses only. In the

[9] It is possible that cases like this are cases of 'mentioning' rather than cases of 'use' and so should have been excluded from the data. Examples like this were included for the sake of completeness, but I would remove them if I were to repeat the study. Regardless, they made up a very small number of tokens and so should not have affected the statistical analyses.

models containing all senses, 'sense' is the most important, but only due to a clear separation between N, COL, and Q as opposed to W and M. Instead, extra-linguistic factors 'author', 'subfield', and 'object' have the most influence. This part of the data analysis points to the CE project being unsuccessful. However, these results are circumscribed by the nature of the corpus and time period studied. Further research is still needed.

8 Discussion

Now that we have a picture of the quantitative and qualitative features of the corpus data, it is possible to look for patterns, conclude how successful this instance of CE was, and evaluate the applicability of the current method to other CE projects in the future.

8.1 Patterns in frequency and senses

Research question (i) asked whether *mass* and *weight* were successfully differentiated by sense, but the frequency data shows that this did not happen. Furthermore, research question (ii) aimed to find time frames and patterns in the diffusion, but this study found that even after 220 years the lemmas were still not used with their corresponding senses. In sum, this data shows that the CE project was not successful, and the question remains as to why.

Since N remains the most prominent sense of *mass*, it is possible that this established usage of *mass* with nominal sense (N), which pre-dates the usage with sense MASS, provides resistance to the allocation of MASS to the same lexical item.

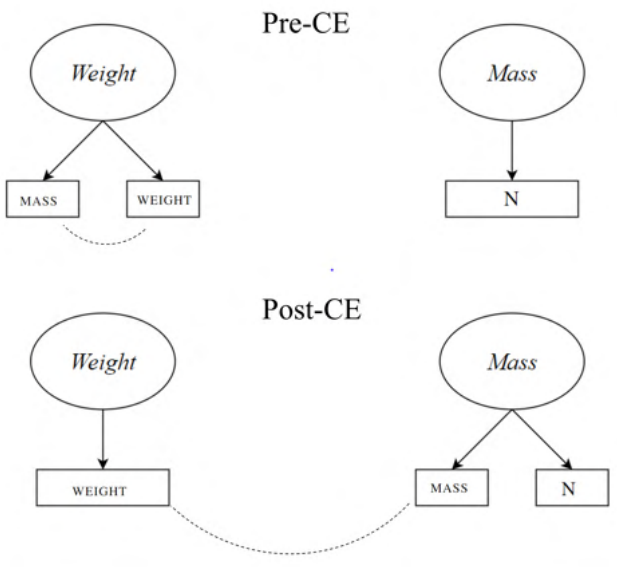

Figure 6 Idealised visual representation of the sense changes.

As can be seen from Figure 6, the word *weight* initially had two related but distinct senses, namely MASS and WEIGHT, in a way that is reminiscent of polysemy. After the re-allocation of sense MASS to *mass*, this led to two unrelated senses, namely MASS and N, belonging to the same lexical item, in a way that is reminiscent of synonymy.

The sense WEIGHT never belonged to the lemma *mass* at any stage in the process, which may explain the one-sided use of *weight* to refer to both MASS and WEIGHT, but *mass* never referring to WEIGHT.

As was mentioned earlier in this work, a project that aims to reduce ambiguity in one area may result in increased ambiguity or polysemy in another place in the lexical network, and this is indeed what we see in the case of *mass*. Synonymy and polysemy are not uncommon and occur frequently in words such as *bank* or *head* but in this case, it has been transferred from one lexical item to another.

Competing senses within the same lexical item may therefore have led to resistance to the CE project. However, it is important to reflect on whether synonymy and polysemy are legitimate blocking forces within CE. Synonymy and polysemy are rife in language, and in most cases, it is not a problem. To investigate this notion more, it is worth looking at another instance of CE.

Some projects, such as Haslanger (2000), choose to keep the lexical item but change the sense associated with it. In this case, the lexical item *woman* remains the same, but the meaning associated with it changes. Other projects such as Scharp (2013) ameliorate a lexical item by moving certain meanings or senses it may have to another lexical item. This lexical item may already exist, as is the case in *mass/weight*, or a new lexical item may be created altogether.

A case of the latter is explained by Baron (2020), who details the evolution of a gender-neutral, inclusive pronoun. Certain new lexical items have been put forward to fulfil this role, including *hasher/himer/hiser, e/es/em* and *ir/e/thon* to name a few. These new lexical items were created because some people were uncomfortable using an existing word with an existing function (such as the use of third-person pronoun *they* as a second-person pronoun) as this new gender-neutral concept. Regardless, *they* has been the most successful of these terms. *They* is also closely related to the previously existing *he/she*, being part of the same pronoun paradigm. This lends doubt to the idea that it is the close relationship between the pre-CE lexical item and post-CE lexical item and sense distribution that creates conflict in the general acceptance of a CE project.

In the case of pronouns, an existing term (*they*) was eventually selected over a new term. However, the same strategy did not work for *mass/weight*, where the transfer of the sense M to a different existing lexical item, *mass*, was unsuccessful. So, appealing to the notions of polysemy and synonymy is not (always) applicable.

Perhaps the answer lies in the communities targeted by the project. The case of pronouns targeted the social domain with the intention of social activism, whereas *mass/weight* targeted the scientific domain to increase precision. These two different aims may lead to different constraints and strategies, but this will also require more research on terminological development in the scientific and vernacular domains. Perhaps it may be the case that synonymy and polysemy are stronger driving

forces in scientific language than in everyday language, but this will require further investigation. The power of synonymy and polysemy to block or inhibit conceptual change will be crucial to future projects of CE.

8.2 Factors in selection

Research question (i) concerned the importance of 'sense' in predicting lemma choice over time. The data shows that in the time-specific models containing all senses, 'sense' stays the highest ranked by the VIR in all periods. However, from inspecting the Ctrees, we know that the splits based on 'sense' are usually separate MASS and WEIGHT from the other annotated senses.

If the CE project is successful, as time goes on the VIRs should have shown an increase in the importance of 'sense' over time in the model containing only the target senses, since this would exclude influence from the other annotated senses. 'Sense' did not necessarily have to become the most important since the change could still be ongoing by the end of the analysis, but it would be expected that it should eventually rank more highly than 'author'. However, this does not occur. Individual choice remains consistently more important than 'sense' in lemma choice. In fact, 'sense' becomes less important over time.

	All senses	Target senses only
1800-1850	Sense, Object, Author	Author, Subfield, Object
1850-1900	Sense, Object, Author	Object, Author, Subfield
1900-1920	Sense, Subfield, Object	Subfield, Author, Object

Table 3 Highest ranked variables in each time period.

So, if not 'sense', what does affect lemma choice? Research question (iii) asked whether extra-linguistic variables have any effect, and the results in Table 3 clearly show that they do. In the models of only the target senses, 'author', 'object', and 'subfield' are consistently found to be the most important predictive factors, consistently ranking above 'sense' in models containing only target senses. We can therefore answer research question (iii) by saying that extra-linguistic factors have an overwhelming effect on lemma choice.

Research question (ii) aimed to look at the time frame of implementation, and diffusion patterns. From Table 3, we see that there is a remarkable consistency in terms of which factors influence lemma choice over time. The same three factors remain the highest-ranked in the models of the target senses only, and 'sense' remains the highest-ranked in the model containing all senses. This suggests there was little change in what factors condition choice.

Lemma choice, therefore, varied as a function of individual variation, the 'subfield' in which the author was writing, and the 'object' about which they are writing. This instance of CE diffused socially and marked discipline boundaries. The hypotheses

stated that if extra-linguistic factors were found to be the most important, then this instance of CE could not be considered successfully implemented.

The most unexpected result from these statistical analyses is the lack of influence of 'year' in the model based on the entire annotated data set. If Newton's CE project was successful, we would expect to see a separation of the senses MASS and WEIGHT over time. This raises the question of time frames within CE projects. When someone engineers a term for amelioration and intends for it to be spread beyond a single individual, they would typically want their new concepts to be successfully spread as quickly as possible since the aim of improving these concepts in the first place is to solve a problem. If the new concept is not used, the problem is not solved. However, we know from linguistic research that semantic change often takes a very long period to diffuse, and it is rare that meaning changes overnight or in a short period. Regardless, 220 years is a long time. Conceptual engineers must therefore be prepared for their proposals to take time to spread.

In the data examined in this study, it is possible that the period looked at was too small to track the change to completion. Only 220 years of data were examined, and linguistic changes may require more time than this to spread. Once the data from 1920 onward becomes available, the study can be extended into the 21st century.

8.3 Overall success

Taking into consideration the results of the data analysis, it appears as if Newton's CE project was not successful. If it were successful, there would have been increasing usage of lemmas with their 'correct' respective senses, and that sense would have been an important conditioning factor in the statistical analysis. Although we see growing structure in how *mass* and *weight* are used in the RSC, this seems to come from the 'subfield', 'author' and 'object'.

However, this leads to more questions. For example, who were the people and subfields that innovated the change? Who helped spread it? What happened after 1920? More work must be done to answer these questions. What we can definitively say is that Newton's proposal was not accepted at large. More polysemy was created, more individual variation emerged, and more division between disciplines developed.

8.4 The future of Conceptual Engineering

Now, it may be tempting to jump to the conclusion that this suggests the outlook for CE is bleak. However, this was only a small study conducted on a single, homogeneous, niche community. Therefore, it is difficult to generalise the results to all future and past instances of CE. Conversely, given this group has a clear common goal of scientific progress, and the case of CE appealed to this goal, the chances of success in a larger and more diverse community may be small. The implementation problem has been frequently discussed by philosophers in a theoretical sense, or in terms of what society would have to be like for it to spread, but to my knowledge, no empirical study has been conducted on an instance of CE until this one. What

was found is that individuality plays a significant role in the usage of engineered concepts and that communities should not be viewed as a homogeneous group that will all accept a certain concept because of a well-argued explanation as to why they should.

8.5 Future avenues of research

This work has found that concepts 'stretch, shrink, or refigure what exactly we are talking about' (Haslanger 2012: 225). This works, and will continue to work, as long as the hearer/reader can identify the intentions of the speaker/writer and infer the correct sense from what is uttered. This notion of speaker and hearer inference brings us to an avenue for possible further research into CE. Semantics and pragmatics have much to offer to the study of CE, both theoretically and experimentally. For example, when a concept is engineered, does it gain the engineered meaning when the speaker starts using it in a new way, or when the hearer starts understanding it in a new way? This is only one of many pragmatic principles that could, and should, be investigated when it comes to CE. Another is the idea of common ground and background knowledge. If the hearer/reader has shared common knowledge with the speaker/writer and knows which topics they are interested in, which theories they support, and what their ideologies are, then this could cause them to infer from the speaker's utterance something different than what is said. For example, in the context of Biology, two conversing biologists may have an implicit understanding that in saying *bodyweight*, they are in actuality referring to the MASS of the human body. This depends on there being this common ground, much like other pragmatic phenomena that have been investigated. This is likely to be a fruitful future endeavour.

This work has focused on scientific concepts, and this meant it has also focused on a niche, more Habermasian (Habermas 1996), group of individuals, in the sense of Habermas' theory of society. While this was necessary for the scope of the current study, there is also the potential that looking at the public more generally, and more social concepts, could result in a vastly different outcome. In a project that aims to target a wider community, more factors could contribute to the uptake of, or resistance to, the ameliorated concept. Many, more diverse, factors could contribute to such a change, including social and political factors that may not be pertinent to a small scientific community. A larger, more varied group leads to even more variation in ideologies and goals, potentially further confounding CE.

9 Conclusion

Newton's CE project had two criteria for success; the consistent usage of *mass* with sense MASS and *weight* with sense WEIGHT, and that the choice of lemma was governed by sense alone. This study found that the case of CE proposed by Newton concerning *mass* and *weight* was unsuccessful. Instead, there was much idiosyncratic variability, and variation based on the subfield and object under discussion.

This is, to my knowledge, the first study of its kind[10] that analyses the implementation of a CE example and as such shows that this work needs to continue to understand how CE (deliberately or not) works to change reality.

The results and discussion in this work have exemplified that linguists have much to offer to CE when it comes to understanding and discussing the implementation problem and that they can contribute without making any normative judgments. By considering sociolinguistic theories and language ideologies in addition to an empirical study, it is possible to evaluate the success of CE projects and contextualise the results of the models. Through similar projects, it may become possible to create metrics to measure the success of such projects and understand which factors need to be considered in the implementation of an engineered term.

[10] Of course, many other works have looked at the results of conscious language change such as in Kenya (Trudell & Piper 2014) or language planning in Belgium (de Groof 2002).

References

Altakhaineh, A. R. M. 2018. The semantic change of positive vs. negative adjectives in Modern English. *Lingua Posnaniensis* 60(2). 25–38. doi:10.2478/linpo-2018-00010.

Andow, J. 2020. Fully experimental conceptual engineering. *Inquiry* 0(0). 1–27. doi:10.1080/0020174X.2020.1850339.

Andow, J. 2021. Conceptual engineering is extremely unlikely to work. So what? *Inquiry* 64(1-2). 212–226. doi:10.1080/0020174X.2020.1850343.

Bacon, F. 1623. *Novum Organum*. P.F Collier.

Bacon, F. 2011. The New Organon. In J. Spedding, R. L. Ellis & D. D. Heath (eds.), *The Works of Francis Bacon: Volume 4: Translations of the Philosophical Works 1*, vol. 4 Cambridge Library Collection - Philosophy, 37–248. Cambridge: Cambridge University Press. Original date: 1620.

Banks, D. 2008. *The Development of Scientific Writing: Linguistic Features and Historical Context*. Equinox.

Baron, D. 2020. *What's your pronoun?: Beyond he and she*. Liveright.

Bauman, R. & C. L. Briggs. 2003. *Voices of Modernity: Language Ideologies and the Politics of Inequality*. Cambridge University Press 1st edn. doi:10.1017/CBO9780511486647.

Breiman, L. 2001. Random Forests. *Machine Learning* 45(1). 5–32. doi:10.1023/A:1010933404324.

Bybee, J. 2006. From Usage to Grammar: The Mind's Response to Repetition. *Language* 82(4). 711–733. Publisher: Linguistic Society of America.

Cappelen, H. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press. doi:10.1093/oso/9780198814719.001.0001.

Cappelen, H. & D. Plunkett. 2020. Introduction: A Guided Tour of Conceptual Engineering and Conceptual Ethics. In *Conceptual Engineering and Conceptual Ethics*, 1–34. Oxford University Press. doi:10.1093/oso/9780198801856.003.0001.

Carnap, R. 1947. *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago, IL: University of Chicago Press.

Carnap, R. 1950. *Logical foundations of probability*. Chicago, IL, US: University of Chicago Press.

Carnap, R. 1962. *Logical Foundations of Probability*. University of Chicago Press.

Clark, A. & D. Chalmers. 1998. The Extended Mind. *Analysis* 58(1). 7–19.

Crookes, W. 1873. VI. researches on the atomic weight of thallium. *Philosophical Transactions of the Royal Society of London* 163. 277–330. doi:10.1098/rstl.1873.0007. https://doi.org/10.1098/rstl.1873.0007.

Cull, M. J. 2020. *Engineering Genders: Pluralism, Trans Identities, and Feminist Philosophy*. Sheffield: University of Sheffield PhD Thesis.

Cull, M. J. 2021. Engineering is not a luxury: Black feminists and logical positivists on conceptual engineering. *Inquiry* 64(1-2). 227–248. doi:10.1080/0020174X.2021.1883476.

Darwin, G. H. 1882. IV. On the stresses caused in the interior of the earth by the weight of continents and mountains. *Philosophical Transactions of the Royal Society of London* 173. 187–230. doi:10.1098/rstl.1882.0005. https://

royalsocietypublishing.org/doi/10.1098/rstl.1882.0005. Publisher: Royal Society.

Davy, J. 1865. Some observations on birds, chiefly relating to their temperature, with supplementary additions on their bones. *Proceedings of the Royal Society of London* 14. 440–457. http://www.jstor.org/stable/112167.

Dawson, H. 2007. *Locke, Language and Early-Modern Philosophy*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511490484.

Desaguliers, J. T. 1728. *The newtonian system of the world, the best model of government an allegorical poem. with a plain and intelligible account of the system of the world, by way of annotations ... to which is added, cambria's complaint against the intercalary day in the leap-year*. A. Campbell.

Deutsch, M. 2020. Speaker's reference, stipulation, and a dilemma for conceptual engineers. *Philosophical Studies* 177(12). 3935–3957. doi:10.1007/s11098-020-01416-z.

Dutilh Novaes, C. & E. Reck. 2017. Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization. *Synthese* 194(1). 195–215. doi:10.1007/s11229-015-0816-z.

Eklund, M. 2014. Replacing Truth? In *Metasemantics*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199669592.003.0012.

Faraday, M. 1840. III. experimental researches in electricity.—seventeenth series. *Philosophical Transactions of the Royal Society of London* 130. 93–127. doi:10.1098/rstl.1840.0004. https://doi.org/10.1098/rstl.1840.0004.

Fischer, S., J. Knappen, K. Menzel & E. Teich. 2020. The Royal Society Corpus 6.0: Providing 300+ Years of Scientific Writing for Humanistic Study. In *Proceedings of the 12th Language Resources and Evaluation Conference*, 794–802. Marseille, France: European Language Resources Association.

Fonteyn, L. & A. Nini. 2020. Individuality in syntactic variation: An investigation of the seventeenth-century gerund alternation. *Cognitive Linguistics* 31(2). 279–308. doi:10.1515/cog-2019-0040. Publisher: De Gruyter Mouton.

Gray, B. & D. Biber. 2018. Academic writing as a locus of grammatical change: The development of phrasal complexity features. In R. J. Whitt (ed.), *Studies in Corpus Linguistics*, vol. 85, 117–146. Amsterdam: John Benjamins Publishing Company. doi:10.1075/scl.85.06gra.

de Groof, J. 2002. Two hundred years of language planning in Belgium. In A. R. Linn & N. McLelland (eds.), *Current Issues in Linguistic Theory*, vol. 235, 117–134. Amsterdam: John Benjamins Publishing Company. doi:10.1075/cilt.235.10deg.

Habermas, J. 1996. *Between facts and norms: Contributions to a discourse theory of law and democracy*. Cambridge, MA: MIT Press. Original date: 1992.

Habgood-Coote, J. 2020. Fake news, conceptual engineering, and linguistic resistance: reply to Pepp, Michaelson and Sterken, and Brown. *Inquiry* 1–29. doi:10.1080/0020174X.2020.1758770.

Hardie, A. 2012. CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17(3). 380–409. doi:10.1075/ijcl.17.3.04har. Publisher: John Benjamins.

Harris, J. 1704. *Lexicon Technicum: Or, An Universal English Dictionary of Arts and Sciences:: Explaining Not Only the Terms of Art, But the Arts Themselves*. Brown, Dan and Goodwin, Tim and Walthoe, John and Newborough, Thomas and Nicholson, John and Benskin, Thomas and Tooke, Benjamin and Midwinter, Dan and Leigh, Thomas and Coggan, Francis.

Haslanger, S. 2000. Gender and Race: (What) Are They? (What) Do We Want Them To Be? *Noûs* 34(1). 31–55. doi:10.1111/0029-4624.00201.

Haslanger, S. 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford University Press.

Hothorn, T., P. Buehlmann, S. Dudoit, A. Molinaro & M. V. D. Laan. 2006a. Survival Ensembles. *Biostatistics* 7(3). 355–373.

Hothorn, T., K. Hornik & A. Zeileis. 2006b. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674.

Jammer, M. 1997. *Concepts of Mass in Classical and Modern Physics*. Courier Corporation.

Jorem, S. 2021. Conceptual engineering and the implementation problem. *Inquiry* 64(1-2). 186–211. doi:10.1080/0020174X.2020.1809514.

Kater, H. 1819. An account of experiments for determining the variation in the length of the pendulum vibrating seconds, at the principal stations of the trigonometrical survey of great britain. *Philosophical Transactions of the Royal Society of London* 109. 337–508. http://www.jstor.org/stable/107509.

Koch, S. 2018. The externalist challenge to conceptual engineering. *Synthese* 198(1). 327–348. doi:10.1007/s11229-018-02007-6. http://link.springer.com/10.1007/s11229-018-02007-6.

Koch, S. 2021. The externalist challenge to conceptual engineering. *Synthese* 198(1). 327–348. doi:10.1007/s11229-018-02007-6.

Koch, S. & J. Ohlhorst. 2022. The Mind-Expanding Powers of Conceptual Engineering. https://www.youtube.com/watch?v=RptQmwlVJOs.

Kopp, H. 1865. III. investigations of the specific heat of solid bodies. *Philosophical Transactions of the Royal Society of London* 155. 71–202. doi:10.1098/rstl.1865.0003. https://doi.org/10.1098/rstl.1865.0003.

Labov, W. 2001. *Principles of linguistic change, Volume 2: Social Factors*. Oxford: Blackwell. OCLC: 47901638.

Labov, W. 2010. *Principles of Linguistic Change, Volume 3: Cognitive and Cultural Factors*. Wiley.

Lakoff, G. & M. Johnson. 1981. *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Locke, J. 1689. *An Essay Concerning Human Understanding*. Thomas Basset.

Locke, J. 1975. An Essay Concerning Human Understanding. In P. Nidditch (ed.), *An Essay Concerning Human Understanding: Clarendon Edition of the Works of John Locke*, Oxford University Press. Original date: 1690.

Losonsky, M. 2021. The History of the Philosophy of Language before Frege. In P. Stalmaszczyk (ed.), *The Cambridge Handbook of the Philosophy of Language*, 51–70. Cambridge University Press 1st edn. doi:10.1017/9781108698283.003.

McDougall-Water, J., N. Moxham & A. Fyfe. 2021. History of the Philosophical Transactions. *https://royalsociety.orgs/journals/publishing-activities/publishing350/history-philosophical-transactions* [Accessed: 16/02/2022].

Meyns, C. 2020. 'Data' in the Royal Society's *Philosophical Transactions* , 1665–1886. *Notes and Records: the Royal Society Journal of the History of Science* 74(3). 507–528. doi:10.1098/rsnr.2019.0024.

Miller. 1857. II. on the construction of the imperial standard pound, and its copies of platinum and on the comparison of the imperial standard pound with the kilogramme des archives. *Proceedings of the Royal Society of London* 8. 87–103. doi:10.1098/rspl.1856.0031. https://doi.org/10.1098/rspl.1856.0031.

"nice n.1". 2022. Oxford English Dictionary. In *OED Online*, Oxford University Press. *http://www.oed.com/view/Entry/126731* [Last accessed 08/04/2022].

"weight n.1". 2022. Oxford English Dictionary. In *OED Online*, Oxford University Press. *https://www.oed.com/view/Entry/26891* [Last accessed 17/02/2022].

"mass n.1". 2022. Oxford English Dictionary. In *OED Online*, Oxford University Press. *https://www.oed.com/view/Entry/114665* [Last accessed 17/02/2022].

"mass n.2". 2022. Oxford English Dictionary. In *OED Online*, Oxford University Press. *https://www.oed.com/view/Entry/114666* [Last accessed 17/02/2022].

Nado, J. 2021. Conceptual engineering via experimental philosophy. *Inquiry* 64(1-2). 76–96. doi:10.1080/0020174X.2019.1667870.

Newton, I. 1687. *Philosophiæ Naturalis Principia Mathematica.* Londini: Apud G. & J. Innys.

Newton, I. 1999. *The Principia: Mathematical Principles of Natural Philosophy.* University of California Press.

Nietzsche, F. 1901. *The Will to Power.* New York: Random House 1st edn.

Pratt, J. H. 1854. On the attraction of the himalaya mountains, and of the elevated region beyond them, upon the plumb-line in india. *Proceedings of the Royal Society of London* 7. 175–182. http://www.jstor.org/stable/111767.

Reed & Stokes. 1871. XVI. on the unequal distribution of weight and support in ships, and its effects in still water, in waves, and in exceptional positions on shore. *Philosophical Transactions of the Royal Society of London* 161. 413–465. doi:10.1098/rstl.1871.0017. https://doi.org/10.1098/rstl.1871.0017.

Richard, M. 2020. The A-project and the B-project. In *Conceptual Engineering and Conceptual Ethics*, 358–378. Oxford University Press. doi:10.1093/oso/9780198801856.003.0017.

Ruíz, R. 1984. Orientations in Language Planning. *NABE Journal* 8(2). doi:10.1080/08855072.1984.10668464.

Scharp, K. 2013. *Replacing Truth.* Oxford University Press.

Schunck, E. 1853. III. on rubian and its products of decomposition. *Philosophical Transactions of the Royal Society of London* 143. 67–107. doi:10.1098/rstl.1853.0003. https://doi.org/10.1098/rstl.1853.0003.

Silverstein, M. 1979. Language structure and linguistic ideology. In P. Cline, C. Hanks & C. Hofbauer (eds.), *The elements: A parasession on linguistic units and levels*, 193–246. Chicago, IL: Chicago Linguistics Society.

Sprat, T. 1667. *The History of the Royal Society.* London: T.R.

Sterken, R. K. 2020. Linguistic Intervention and Transformative Communicative Disruptions. In *Conceptual Engineering and Conceptual Ethics*, 417–434. Oxford University Press. doi:10.1093/oso/9780198801856.003.0020.

Trudell, B. & B. Piper. 2014. Whatever the law says: language policy implementation and early-grade literacy achievement in Kenya. *Curr. Issu. Lang. Plan.* 15(1). 4–21.

Woolard, K. A. & B. B. Schieffelin. 1994. Language Ideology. *Annual Review of Anthropology* 23(1). 55–82. doi:10.1146/annurev.an.23.100194.000415. Publisher: Annual Reviews.

Appendices

A.1

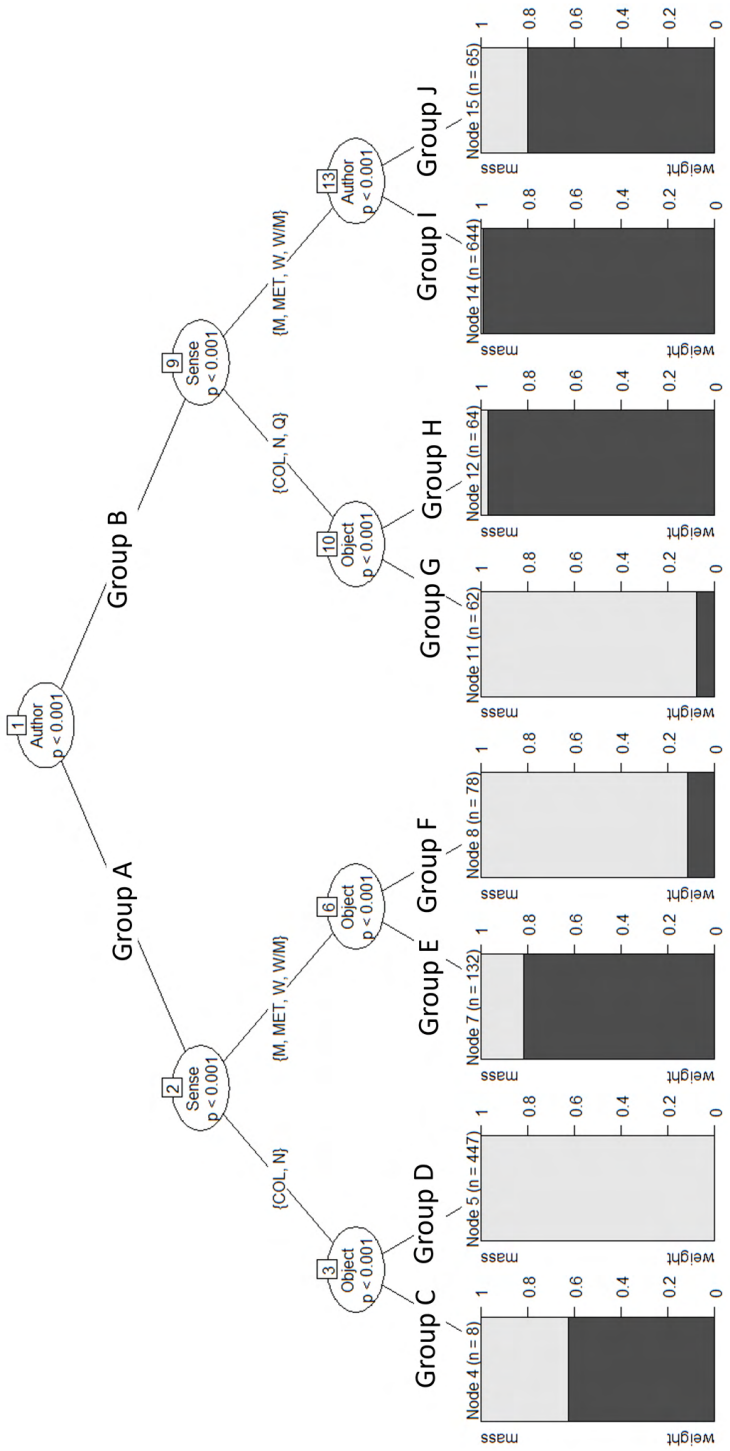

Figure 7 Simplified Conditional Inference Tree of all annotated data

Group	Split Based on	
A	Author	A.W.Hofmann, C.Hatchett, Griffiths & Glazebrook, F.A.Abel, F.O.Bower, G.B.Airy, G.H.Darwin, G.Pearson, H.Hennessy, H.Watney, J.D.Forbes, J.H.Jeans, J.H.Pratt, J.Ivory, J.Lockheart, J.L.Clarke, J.N.Lockyer, J.Tyndall, L.S.Beale, M.Barry, M.Faraday, R.Mallet, T.Anderson, W.B.Hardy, W.C.Williams, W.Hopkins, W.K.Parker, W.Thomson
B	Author	A.E.H.Tutton, B.Count, B.C.Brodie, B.Thompson, C.T.Heycock, D.Mendel, D.Thoday, E.Hodgkinson, Halnan, Marshall, Yule &Langley, E.J.Reed, Reed & Stokes, E.Sabine, F.Baily, G.Dreyer, Matthaei & Darwin, G.Rennie, G.Shuckburgh, H.Cavendish, H.Davy, H.E.Roscoe, H.Kater, H.Moseley, H.Tomlinson, J.Barnard, J.Davy, Gray, Henderson & Kelvin, J.Herschel, J.Joly, Joly & Fitzgerald, Lawes & Gilbert, Manley & Poynting J.P.Joule, J.T.Desaguliers, J.W.Mallet, M.Raper, R.Kirwan, T.Graham, T.Thomson, Thorpe & Rodger, W.Crookes, W.Fairbairns, W.H.Miller, W.Marcet
C	Object	GEN, STAN, WO
D	Object	BIO, C, EQ, F, GEO, MAT, MATTER, S
E	Object	ABS, BIO, BO, C, CONCEPT, EQ, FOR, GEO, MAT, STAN, WO
F	Object	F, MATTER, S
G	Object	BIO, C, EQ, F, GEO, MAT, MATTER, S, VM
H	Object	CONCEPT, STAN, WO
I	Author	A.E.H.Tutton, B.Count, B.C.Brodie, B.Thompson, C.T.Heycock, D.Mendel, D.Thoday, E.Hodgkinson, Halnan, Marshall, Yule & Langley, E.J.Reed, Reed & Stokes, E.Sabine, F.Baily, G.Dreyer, Matthaei & Darwin, G.Rennie, G.Shuckburgh, H.Cavendish, H.E.Roscoe, H.Kater, J.Barnard, J.Davy, Gray, Henderson & Kelvin, Joly & Fitzgerald, Lawes & Gilbert, J.P.Joule, J.T.Desaguliers, J.W.Mallet, M.Raper, R.Kirwan, T.Graham, T.Thomson, Thorpe & Rodger, W.Crookes, W.Fairbairs, W.H.Miller, W.Marcet
J	Author	H.Darwin, H.Moselet, H.Tomlinson, J.Joly, Manley & Poynting

Figure 8 Group data for the abbreviated Ctree of all data for all senses in Figure 7

A.2

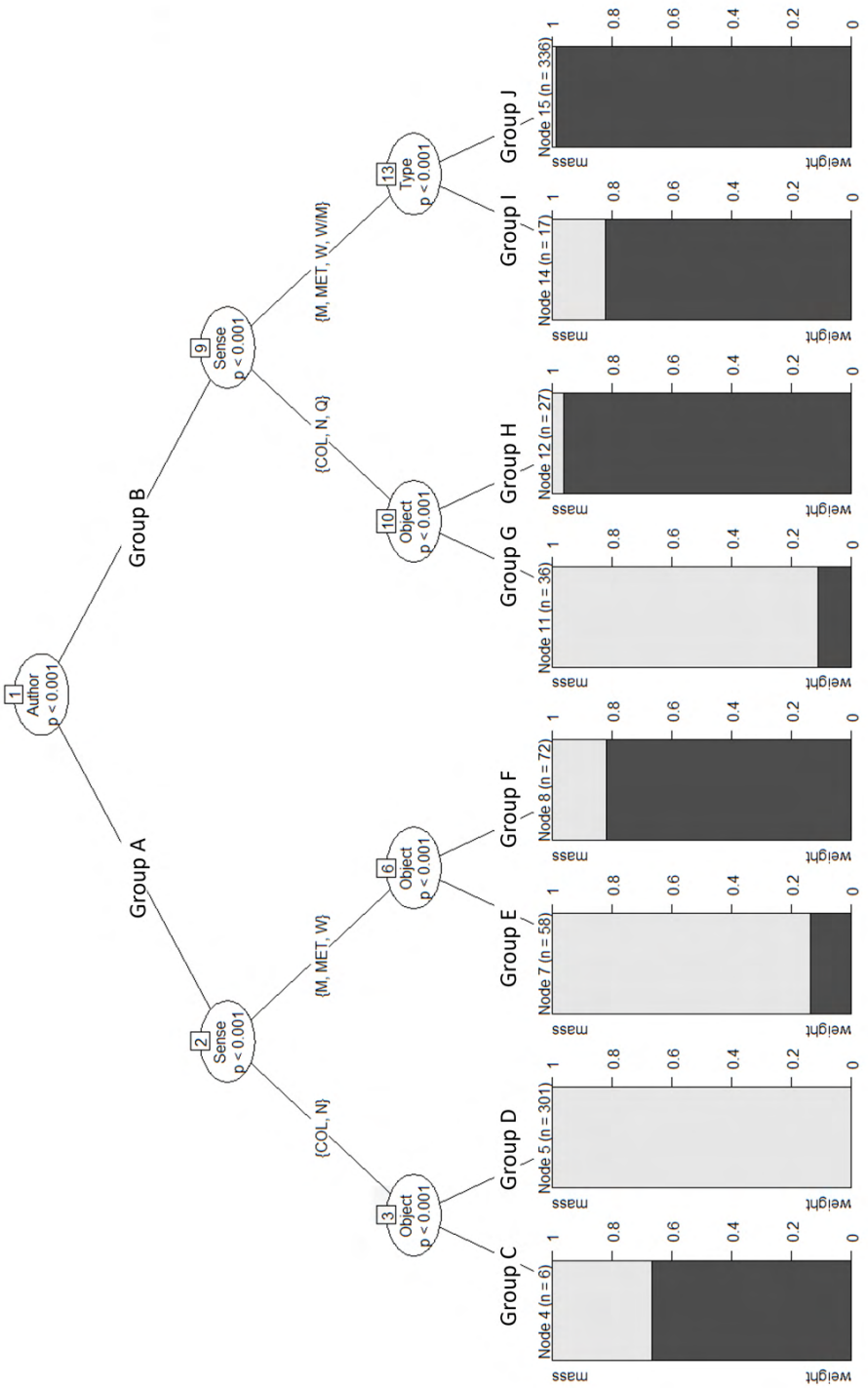

Figure 9 Ctree of all senses between 1850-1900

Group	Split Based on	
A	Author	A.W.Hofmann, Griffiths & Glazebrook, F.A.Abel, F.O.Bower, G.H.Darwin, H.Henessy, H.Watney, J.H.Pratt, J.Lockhart, J.L.Clarke, J.N.Lockyer, J.Tyndall, L.S.Beale, M.Faraday, R.Mallet, W.C.Williamson, W.Hopkins, W.K.Parker, W.Thomson
B	Author	B.C.Brodie, C.T.Heycock, D.Mendel, E.Hodgekinson, E.J.Reed, Reed & Stokes, E.Sabine, G.B.Airy, H.E.Roscoe, H.Kopp, H.Mallet, H.Tomlinson, J.Barnard, J.Davy, Grey, Henderson & Kelvin, J.Joly, Joly & Fitzgerald Lawes & Gilbert, J.P.Joule, J.W.Mallet, T.Graham, Thorpe & Rodger, W.Crookes, W.Fairbairn, W.H.Moorby, W.Marcet
C	Object	EQ, F, S
D	Object	ABS, BIO, BO, C, FOR, GEO, MAT, WO
E	Object	BIO, C, EQ, F, GEO, MAT, S, VM
F	Object	CONCEPT, STAN, WO
G	Object	GEN, MATTER, WO
H	Object	BIO, C, EQ, F, GEO, MAT, S
I	Type	experiment, lecture
J	Type	abs, abstract, article, fla

Figure 10 Group data for Ctree all senses 1850-1900 in Figure 9

A.3

Code, data, and all figures are available in the following link:
https://drive.google.com/drive/folders/1SZ_2kniQVyGodgPf2fdVSSd1rlnGd2pv?usp=sharing

Nina Haket
The University of Cambridge
nch35@cam.ac.uk