Concepts and their Engineering

Heimir Geirsson

Iowa State University

Conceptual engineering is broadly concerned with the assessment and improvement of concepts, where concepts are thought of as representational devices. Even a cursory survey of the literature shows that there seems to be little agreement about what is involved in conceptual engineering or whether it is even possible. On some accounts, what we consider conceptual engineering is us trying to clarify the meaning of the concept, i.e., trying to uncover its true meaning. That does not seem to involve conceptual engineering as the *concept* is not being changed or modified. On other accounts conceptual engineering consists of us changing our understanding of the world, with the eventual conceptual change reflecting that. Some claim that conceptual engineering consists of finding out what the concepts *should* mean, resulting in either in providing an ameliorative account of the concept or, perhaps, a revisionary account of it. What these accounts have in common is that they all are working with concepts. However, they typically do not clarify what kinds of concepts they are working with. It is instructive to provide a very brief survey of types of concepts with an eye on whether the various accounts of conceptual engineering are too narrowly focused, i.e., whether the accounts apply to only one type of concept, or whether the accounts apply more generally. Ideally, if one provides an account of conceptual engineering it should apply to concepts across the board, assuming that such an account is possible. If an account that applies across the board is not possible, it becomes incumbent on the conceptual engineer to be very clear on what type of concept she is working with.

The importance of clarifying what kind of a concept we are working with will become clear once we present what is often thought of as a key challenge to any kind of conceptual engineering; namely what Herman Cappelen calls the Strawsonian Challenge:

> "To offer formal explanations of key terms of scientific theories to one who seeks philosophical illumination of essential concepts of non-scientific discourse, is to do something utterly irrelevant--is a sheer misunderstanding, like offering a textbook on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart… Typical philosophical problems about the concepts used in nonscientific discourse cannot be solved by laying down the rules of exact and fruitful concepts in science. To do this last is not to solve the typical philosophical problem, *but to change the subject*." (Strawson 1963, 505)

In the context of conceptual engineering, the above is generally understood as a question about topic continuity. Namely, if the meaning of a concept changes, then there is a danger of the subject changing with it. Sally Haslanger (2000, 34) makes the same point as follows: "Revisionary projects are in danger of providing answers to questions that weren't being asked." Generally, the worry that the challenge raises in the context of conceptual engineering is twofold: first, that what is being discussed changes, i.e., that the topic of discussion changes, and second, that the discussants start talking past each other as they are talking about different things.

Discussion of the Strawsonian Challenge, it appears to me, has focused primarily on one specific type of concept, namely the classical account of concepts. The meaning of a concept according to the classical account is typically specified with necessary and sufficient conditions. A change of those specifications is likely to lead to, for example, changes in the extension of the concept. And if topic continuity is dependent upon meaning determining extension, then a change in meaning with the

corresponding change in extension invites the Strawsonian Challenge. As regards Haslanger's worry, it depends on the extent of the revision whether there is a danger of one answering questions that weren't being asked.

The reason for my somewhat skeptical attitude towards the seriousness of the Strawsonian Challenge is that it is very common that we communicate with each other and have a fruitful discussion about what we regard as a well-focused topic without us sharing the exact same understanding of the key concepts being employed. Unlike the classical account of concepts, accounts of other types of concepts make it unlikely that any two speakers share exactly the same concept. The next section, which provides a brief account of various types of concepts, will explain my point, and make clear that the meaning of our concepts often differs from one person to the next as well as over time. The second section will discuss what it is that secures topic continuity given differences in meaning. Building on Putnam's account, I will argue that it is reference construed as a *same-as* relation in the case of natural kinds. The third section will focus on socially constructed kinds where the *same-as* relation may be grounded in properties that are not essential to the kind, but rather properties that we use as key markers for our social constructs. These may be, for example, morphological differences that we use to demarcate races. The fourth section will further discuss the hard and soft boundaries provided by the essential properties of natural kinds vs the non-essential properties of socially constructed kinds. The latter helps explain the possibility of both unintended drifting in meaning as well as the intentional changes brought about by conceptual engineering. Depending on the type of concept that we are working with, conceptual engineering is in many cases best carried out in social practice, providing examples that challenge and eventually help modify our representations of the relevant kinds.

1. **Which concepts?**

On **the classical account of concepts,** concepts are viewed as something like Platonic forms or Fregean senses. The former reside in the world of Forms, only accessible to select souls, while the latter reside in the "third realm." Both Platonic forms and Fregean senses are fixed, while Frege suggests that opinions about the sense of a given term might differ when it comes to natural languages. Consider the word 'bachelor'. The word has a meaning that expresses a concept. In this particular case we believe we have a pretty good grasp of the concept, namely that the concept is captured by the meaning of "an unmarried male." The concept is captured by necessary and sufficient conditions.

However, not all cases are this clear and simple. Consider 'knowledge'. On the classical account the term expresses a concept captured by necessary and sufficient conditions and we have spent countless hours, days, and years trying to become more clear about the concept. That is, we have been involved with *conceptual analysis* rather than conceptual engineering when trying to clarify the meaning of 'knowledge'. Each account we provide is meant to capture the true meaning of the concept -- not change it. If we assume that conceptual engineering consists of modifying or altering concepts, then conceptual analysis, generally, does not involve conceptual engineering. Instead, it involves our attempts to better grasp the concepts.

At the level of words, it is often assumed that 'concept' and 'linguistic meaning' are synonyms. If one changes the meaning of a word, then one has changed the concept. Conceptual engineering, then, consists of changing the meaning of words by either redefining them or revising the meaning. In some cases, revisionism replaces conceptual analysis. Consider Haslanger's treatment of *woman*, *race*, *gender*, and related concepts. Haslanger argues that our task is not simply to explicate a given concept and/or discover what object we normally take to fall under the concept. Instead, we should

ask what purpose is served in having the concept, whether that purpose is well conceived, and what concept would serve us best given the purpose (Haslanger 1999; 2012; 2000). In the case of the concept *woman* Haslanger argues that an important purpose of the concept is to subordinate people on the basis of female properties. We should fight against such subordination. Her suggestion is to redefine the concept so that the implicit negative purpose is built into the concept. Once we do that, we see clearly what the concept is used for, thus allowing us more clearly to fight against the given purpose. So, a part of Haslanger's definition of 'woman' is systematic subordination on the basis of female properties (Haslanger 2000, 39). More recently, Haslanger explains that "…by appropriating the terms 'woman' and 'man,' I problematically excluded some women from being counted as women and some men from being counted as men. Although my view does *not* require that one have male genitalia to be a man or female genitalia to be a woman, it does require being subject to subordination/privileged that is linked by ideology to the local bodily markers of reproductive role." (Haslanger 2020, 236)[1]

Is Haslanger simply replacing one concept with another, assuming that she is working on a concept? Many would argue that she is doing just that. If I told someone that in Norway (a bastion of equal rights), unlike in the U.S.A., there are many more adult females than there are adult women, then it is likely that my audience, most of whom have a pretty good grasp of the concepts *female* and *woman* and are competent users of 'females' and 'women', would be perplexed. Haslanger's definition entails sufficient change in the concepts' extensions that it is very likely that speakers using the concepts will talk past each other. Still, one can make a case for Haslanger working on the very same issue as the epistemologist who is working on knowledge. Namely, she is clarifying concepts, getting

---

[1] Here Haslanger states that she is not a Fregean about concepts and that instead of engineering concepts she is appropriating terms. However, whether she is engineering the concept woman or appropriating the term woman, the end result is that she is redefining the term.

closer to their true meaning, and that meaning can be stated in terms of necessary and sufficient conditions.

While the classical account of concepts relies on unchanging entities that we are trying to grasp with necessary and sufficient conditions, **the prototype account of concepts** does no such thing. Instead, the prototype account is built on the idea that concepts are prototypes, which in turn consist of a body of knowledge about the likely properties that members of the class that fall under the concept possess or, alternatively, consists of what one considers to be the most typical member of a category.[2] The relevant knowledge of properties might focus on certain properties being instantiated (e.g., having feet), or properties being instantiated to a degree (e.g., being tall, being sweet). A feature of the account is that most prototype properties are neither necessary nor sufficient for an object being classified as falling under the concept. My prototype of a dog might include barking and four legs, and that doesn't prevent me from classifying Fido, a non-barking three-legged creature, as a dog provided that other properties appropriately fit the prototype. While objects might resemble a prototype to a degree, not all properties have equal weight. Some properties are generally considered more typical for the type than other properties and so receive greater importance. It might be helpful to think of the properties assigned to the prototype as lodged in a web, where the more important properties are centered and less important properties farther from the center.

Putnam flirted with a prototype account of concepts when he introduced his twin-earth thought experiment and used it to motivate an account of meaning. There he writes that the meaning of a natural kind term is complex. In the case of 'water' the meaning includes *syntactic markers*, such as the name being a mass noun, *semantic markers*, such as water being a natural kind and a liquid,

---

[2] Rosch and Mervis (1975) favor the latter of the two.

stereotypes (prototypes), such as water being colorless, tasteless, and finally its extension, namely H_2O (Putnam 1975, 269).

While Putnam flirts with a prototype account of concepts, it is clear that Haslanger is not doing that. She states her conditions for being a woman in terms of necessary and sufficient conditions, while a prototype account focuses on degrees of similarity. With Haslanger's necessary and sufficient conditions, one either is or is not a woman. On the prototype account, one might be judged to resemble a woman to a degree, depending on how similar one is to the prototype. That is certainly a result that Haslanger does not and should not welcome.

The prototype account is open to a *concept* being changed. As we have made discoveries about water, we have adjusted what we might consider a good prototype. For example, while our representations of water have retained the macroscopic properties of water, we have added various microscopic properties. Similarly, while our initial prototypes of tigers likely included the animal being yellow and black striped, the discovery of albino tigers changes the prototype and the likelihood of the animal being white and black.

But, the objection goes, the fact that prototypes can change is consistent with concepts being classical. The kind of change involved would then be a change in conceptions rather than a change in concepts, where conceptions are seen as accompanying elements to concepts. Our conceptions might differ from the concepts due to various reasons, including that we only have a partial grasp of the relevant concept or that while one might possess the concept one does not possess the application conditions of the concept. However, note that this reply treats the accompanying conception, and not the concept, as the main representative device. Our representations change because our conceptions change; it is our conceptions that are providing the relevant explanations and thus doing the heavy lifting.

The prototype theorist, on the other hand, can easily acknowledge that prototypes are nothing but conceptions and so when our conceptions change the prototype concepts change. Same applies to the exemplar account of concepts (below).

The **exemplar account of concepts** is built on the idea that concepts consist of exemplars. The account suggests that we create, for example, a "dog category" by storing in our memory a collection or representations of the dogs that we have encountered, or at least a representative sample of them. Prototype and exemplar accounts are similar in that both assume that when one categorizes an object then one does so on the basis of resemblance. A significant difference is that in the case of the prototype account, one categorizes based on resemblance to a single prototype, while in the case of the exemplar account one can compare with multiple exemplars.

Consider, for example, the concept *infrastructure*. For some time, we have thought that it included roads, bridges, and transportation. The discussion in the U.S. senate about the 2021 infrastructure bill now also recognizes that broadband is a part of the concept. There are debates about what else the concept includes or should include. Climate resilience actions of various kinds? Pre-K-12 support for parents? Perhaps, but not clearly so. But we do not have to go very far back in time to see that broadband would not have been a part of the discussion of what is included in the concept. It, clearly, is a new addition to the meaning of 'infrastructure' and including broadband changes the concept *infrastructure*. And what counts as broadband is somewhat unspecified. While we agree that it is high-speed internet, the physical implementation differs depending on the technology used. Given this, one can understand the concept infrastructure to be an exemplar concept, where politicians in

the U.S. house and senate are debating what should go into the concept as proper exemplars. In a way, the politicians are engaged in conceptual engineering.[3]

Finally, some philosophers advocate externalism when it comes to content, i.e., the content of concepts is determined, at least partly, by the world. One is most likely to encounter this view where objects are thought to constitute a natural kind. Thus, we might have the natural kind tigers, and the concept *tiger*. As we gather information about tigers our understanding of the concept might improve. Accordingly, some argue that conceptual engineering, when it comes to these concepts, consists not in a change of concepts, but rather in acquiring additional information about the world that allows us to better understand the concept.[4]

Why so many accounts of concepts? The account that philosophers have focused on through the centuries, the classical account, certainly has its place. It is very useful, for example, to provide necessary and sufficient conditions for various legal, medical, and theoretical terms. When appropriate, such meaning specification provides unparallel clarity. What, for example, is the difference between a killing, a murder, and letting someone die? A clear definition of each can spell that out. It should be clear, though, that even though scholars use clear and precise definitions, it does not follow that laypeople do the same. A mathematician knows the definition of a circle; a definition that is very simple by mathematical standards. Very few laypeople are likely to know that definition even though they can think about circles and are reliably able to recognize circles. Unlike

---

[3] There are of course additional accounts of concepts. For example, **The theory theory account**, for example, is primarily advocated and used by developmental psychologists who claim that the body of knowledge in a concept is there to explain *why* something is as it is. Accordingly, they view the relevant concept as consisting of various laws and functional properties. Amy Thomasson (2020) argues for a functional account.

[4] Sara Sawyer (2018; 2020a; 2020b; 2020c) develops an account of concepts that recognizes kinds as concepts.

mathematicians, it is not likely that lay people possess the classical concept of circle. Instead, when classifying a shape as a circle, they are more likely to rely on prototypes and/or exemplars.

Experimental psychologists tend to focus on categorization of objects when they develop their notions of concepts.[5] What is it that enables me, for example, to quickly categorize Pepper as a dog? When it comes to tasks such as categorization, empirical psychologists find the classical account of concepts rather useless. Instead, the prototype account and the exemplar account provide much better explanation of how such categorization takes place, even though the explanations that each account provides disagrees with the other.

What this overview of concepts makes clear is that it is very unlikely that any two people have the exact same prototype and exemplar concepts. It is very unlikely that two people have the same prototype concept *dog* because they have had different encounters with dogs and most likely include different properties in the prototype. It is equally unlikely that any two people have the same exemplars in their dog concept. And it is very likely that people have incomplete grasps of many, if not most classical concepts. Clear, precise, and correct definitions of such concepts are hard to come by. In spite of having different understandings of, for example, the concept *dog*, we are fully capable of having meaningful discussions about dogs.

In light of the above, the idea underlying the Strawsonian Challenge, namely that people who don't share a concept are in danger of not talking about the same thing or talking past each other, seems strange. The challenge applies to one type of concept, but clearly not to all other types. As long as we are talking about classical-style concepts where meaning is given by necessary and sufficient conditions, then a change in the conditions specified will in most cases lead to a new meaning and a

---

[5] For a discussion on various empirical work done by psychologists as they develop their notions of concepts see Machery (2009).

new classification of objects that fall under the concept. In some cases, a seemingly simple change in a definition can alter its meaning significantly, or significantly change what falls under the concept. However, when it comes to other types of concepts, such as prototype concepts and exemplar concepts, then it is clear that the concepts of the same object/kind differ from one person to the next, and in spite of that people communicate successfully. For example, people who have different prototypes or different exemplars of dogs can successfully talk about dogs, and people who have different prototypes or exemplars of fruit can successfully talk about fruit. They can, for example, discuss without any confusion about what they are talking about the pros and cons of including fruit in their diet, or whether dogs make good pets. And politicians who have different exemplars in their concept *infrastructure* can successfully discuss and debate whether and how they should fund infrastructure projects. The question then becomes; what is it that allows people who do not share a particular concept to nevertheless converse and talk about what the concept is of? And since we are talking about concepts as representational devices the question, more plainly stated, is as follows; what is it that allows people who don't share the same representations to nevertheless converse and talk about what is represented? It therefore turns out that the Strawsonian Challenge, when applied, for example, to prototype and exemplar concepts, is a very general question about how we can talk and think about the same objects or kinds while representing them differently.

2. **Natural kind examples**

Let us look at Putnam's twin earth thought experiment as an example of how reference is secured. I, like Putnam, accept direct reference and externalism.

In his well-known twin-earth thought experiment Putnam assumes that twin-earth is like earth in all respects except for one. While the liquid that runs in streams and fills lakes and oceans on earth is

H_2O, the liquid that runs in streams and fills lakes and oceans on twin-earth has a different and complicated chemical formula that Putnam abbreviates as *XYZ*. In spite of having a very different chemical composition from water, what twin earthers call 'water', namely *XYZ*, looks like water on Earth and, at least to the layperson, has the properties that water does on Earth. That is, the known phenomenal properties of water on earth and twin-earth water, or twater, are the same.

One of the main goals Putnam had in his article was to show that there is an external element to meaning. He asks us to consider $Oscar_1$, who is a typical earthian speaker of English who does not know the chemical composition of water, and $Oscar_2$, a typical twin-earthian speaker of twin-English who does not know the chemical composition of twater. The two Oscars are duplicates in feelings, thoughts, inner monologue, etc. In spite of that, since the extension of 'water' differs on earth and twin-earth, the meaning of 'water' for $Oscar_1$ is different than it is for $Oscar_2$. Even if the two are in the same psychological state, when $Oscar_1$ thinks water thoughts, his thoughts are about a different substance than when $Oscar_2$ thinks water thoughts.

Given that many followers of the direct reference account of names have argued that the meaning of names of individual objects is their referent, Putnam's use of the words 'means' and 'meaning' is unfortunate, as it sometimes suggests that Putnam is claiming that the meaning of a kind term it its referent. But that is not so. Other factors constitute parts of the meaning of kind terms on Putnam's account. In the case of water, the meaning includes *syntactic markers*, such as the name being a mass noun, *semantic markers*, such as water being a natural kind and a liquid, *stereotypes*, such as water being colorless, tasteless, and finally its *extension*, namely H_2O (Putnam 1975, 269).

Given the complex meaning of 'water', how can we ground our reference such that the term refers to a given kind and not some other kind? Putnam provides a sketch of an answer.

Suppose I point to a glass of water and say 'this liquid is called water' (or 'this is called water, if the marker 'liquid' is clear from the context). My 'ostensive definition' of water has the following empirical presupposition: that the body of liquid I am pointing to bears a certain sameness relation (say, *x is the same liquid as y,* or *x is the same$_L$ as y*) to most of the stuff I and other speakers in my linguistic community have on other occasions called 'water'. (Putnam 1975, 269)

A key to successful grounding is that the empirical presupposition is true. If I mistakenly point to a glass full of gin when trying to provide an ostensive definition of 'water', then I do not intend the definition to be accepted.[6]

Once reference has been fixed and the resulting name is passed on, people on earth who use 'water' refer to H_2O. People on twin-earth, on the other hand, who use 'water' are referring to *XYZ*. Were an earther to travel to twin-earth and casually strike up a conversation about water once she steps out of the transport vehicle, then she and her twin-earth respondent would be talking about two different things. However, note that on this account it makes no difference whether those who use 'water' know about the respective molecular structure of water. The term refers in virtue of its sameness relation, the *same-as* relation, together with the user's intentions to use the name with the same reference as it had when they acquired it, and not in virtue of the details of the beliefs that people have about water. Users of the name during the 17th century referred, respectively, to H_2O and *XYZ*, as do people who use the term during the 21st century.

---

[6] So what is the difference between a Putnam-inspired *same as* relation and Cappelen's samesaying (Cappelen 2018)? Very briefly, Cappelen's samesaying is grounded in meaning in context, whereas the *same as* relation is grounded in reference. At the same time, the *same as* relation includes, in addition to the reference relation, meaning elements that are also a part of the meaning of a term. Note, though, that on Putnam's view one can work on and refine what the *same as* relation relates us to. For example, a water sample is dirty and impure, and scientists can further specify exactly what it is that counts as water.

In spite of the reference of 'water' being retained over the centuries the meaning, in a broad sense, has changed. The various users of 'water' on earth are now likely to know about its chemical composition and so *that water is H_2O* is now likely a part of the paradigm concept *water*. The information might be front and center for some, such as a chemist, or it might be in the background, as for most people who don't care much about chemical features but more about our daily use of water. However, the change in meaning does not affect reference. Reference is secured by a sameness relation and the speakers' intent to use the term to talk about the clear liquid.

Similarly, people are likely to have variations in the concept depending on how they are acquainted with water. The prototype, after all, is likely to represent the features that each person is familiar with, where some of the features that each person tends to focus on are more prominent than the rest. Thus, a 17th century Icelander is likely to have "freezes when cold," "abundant in mountain streams," and "is present in geysers" as a part of her prototype of water. A 17th century person who lives in a desert region close to the equator is not likely to have any of these as a part of their concept. Nevertheless, were they to meet, they would be able to have a conversation about water. What is more, as they converse, each is likely to experience a change in her respective concept of water as each gains additional information about water.[7]

Notice the difference between Oscar[1] traveling to twin-earth and striking up a conversation once he steps out of vehicle, and me traveling back in time to the 17th century and striking up a conversation with a local about water. Unlike the two Oscars, the two of us would be talking about the same

---

[7] Compare this case and the conceptual change with Steffen Koch's proposal for conceptual change for the kind externalist (2021). Koch goes to some length to explain how conceptual change can be intentionally affected by telling a story about how reference change can gradually take place. When the reference change has taken place, then meaning has changed. However, Koch's discussion ignores the point that for kind externalists there is more to meaning than reference. In addition to the kind referred to the subject also has various representations of the kind, and those constitute a part of the kind terms meaning. The fact that there is more to meaning than what is referred to allows one to take steps to change one's representation of the kind without changing the reference of the term.

thing. It is true, of course, that my concept differs from that of the 17th century person since my representations of water include many uses of it that were not available and/or familiar to a 17th century person. However, the reference is not determined by those meaning elements. Over time and with increased knowledge the meaning of 'water' has changed, but the subject matter has not changed. That the meaning has changed with increased knowledge should not come as a surprise. But remember, in many cases the representation, the prototype, that each person has at any given time likely differs from that of other people. That is consistent with the nature of prototype concepts.

Consider a different example, courtesy of Sainsbury. Assume, as he does, that at one time people thought that whales were fish and that they used the term 'whale' accordingly, talking about them as a type of fish. Now, of course, people think of whales as mammals and not as fish. This, Sainsbury claims, gives rise to a dilemma. Either the sentence "whales are fish" means the same in ancient times or it does not. If we say it does, then that fails to accommodate the fact that different uses clearly determine different meanings. If we say it doesn't, then that fails to accommodate the fact that there is substantive disagreement across the two times and so the topic remains the same, namely whales (Sainsbury 2014).

Sarah Sawyer argues that "The diagnosis of the puzzle lies in the recognition that if linguistic meaning is the only representational element in our theory, it is subject to inconsistent constraints. Linguistic meaning cannot both supervene on use *and* determine a stable subject matter….The solution to the puzzle, then, begins with the recognition that two representational elements are required; one to supervene on use, and the other to determine a stable subject matter." (Sawyer 2020b, 386-7) Sawyer, an externalist Fregean, argues that the latter is the function of (classically understood) concepts. These concepts are understood as non-descriptive components of thought,

and they are individuated by relations to objective properties. Hence, while the linguistic meaning of the word 'whale' has changed over time, the term expresses the same concept at the two times and hence concerns the same subject matter (Sawyer 2020b, 387). While people's conceptions of objects might change, Sawyer argues, a concept has its representational properties essentially (Sawyer 2020c). Sainsbury's example, Sawyer argues, forces the introduction of classical concepts.

There is, however, a simpler way to explain how the meaning of 'whale' changes while the subject matter remains stable over time, namely simpler in the sense that it does not introduce a classical concept. One can adopt prototype or exemplar accounts of concepts, explain how additional and/or different exemplars or modifications of the prototype can change meaning over time, and at the same time explain that the term 'whale' refers, Putnam style, to whales in virtue of a sameness relation to the kind that was used when the reference of the name was fixed, thus securing the same subject matter when the name is used with the appropriate intension. This alternate explanation, unlike Sawyer's account, does not force the introduction of classical concepts with their essential properties.

Prototype concepts and exemplar concepts have all the earmarks of what Sawyer calls conceptions; namely, they can and do vary from one individual to the next as well as over time. There is no need to insist that these concepts exist apart from the beings that entertain them. They are, in effect, concrete particulars; individual representations of objects and kinds.[8]

Sawyer's stable concepts reflect a view that concepts represent a metaphysical reality; a view that there are natural kinds with fairly clear boundaries. However, Foster-Hanson and Rhodes point out that such a view is at odds with modern views of the structure of the world (2021). As they point

---

[8] The representations of non-classical concepts are not limited to linguistic meaning. Instead, these concepts are open to a variety of non-linguistic representations, including pictorial representations. There is more to meaning than linguistic meaning.

out, despite people having intuitions about animal species being stable and having immutable properties, nature allows for animals that are "sort of" members of one kind or another, such as a hybrid grizzly and polar bear. And whether people represent race as a natural kind in the sense that it is an objective way of categorizing people depends on multiple factors, including political views, the diversity of the local environment, and their own membership in a majority or minority group. As Foster-Hanson and Rhodes argue, while natural kinds *may* exist in the world, natural kind concepts only exist in the mind.

3. **Stable and Unstable Kinds**

I do not doubt that natural kinds exist. However, it is very likely that there are fewer natural kinds than what we have assumed. Water, whale, woman, gender, and Hispanic have all been treated as natural kinds in the past. Nowadays only the first two are treated as clear natural kinds; the remaining three are generally viewed as, or at least more commonly viewed as socially constructed kinds. The first two are likely to be thought of as categories that correctly cut nature at its joints; the latter three are to an ever-greater extent thought of as categories that are created by humans. However, go back a few decades and people were likely to think of the latter three as correctly cutting nature at its joints as well. That is, they were wrongly thought of as clear examples of natural kinds.

What determines whether we think of a kind as a natural kind is not the kind itself. Instead, as Foster-Hanson and Rhodes argue, whether a kind is viewed as a natural kind is likely determined through a combination of other factors, including psychological, linguistic, and cultural factors. They point out that young children look to the experts in their community (adults) to learn how to divide the world up into meaningful units of experience. Language plays a big role in how this is

accomplished. Generic statements (e.g., "birds lay eggs") describe abstract categories and very young (thirty months) children understand generics to describe such abstract categories instead of particular examples ("this bird lays eggs"). Through this process children, within the first few years of life, start to generalize information about abstract categories, and they start to form beliefs about members of the categories having common properties that likely reflect intrinsic and common causes.[9] The process that Foster-Hanson and Rhodes describe has the advantage of accounting in a reasonable way for cultural variability of what we might count as a natural kind, as well as variability over time of what we might take to be a natural kind.

Biggs and Dosanjh develop a view that explains why we are likely to assume that some of our concepts are of natural kinds. In short, they argue that "…concepts develop to conform to our demand for explanations that are *simple and powerful* (roughly, as simple as they can be without losing power)…The theory that the samples have a single common hidden structure is exceptionally simple and powerful. It appeals to just one entity to explain clustering." (Biggs and Dosanjh 2021, 312)

What the two explanatory accounts above have in common is that they explain how we are driven to assume that there are very stable kinds of the sorts that we call natural kinds. The two explanations imply that it is quite likely that some of the items that we group together as natural kinds are not at all natural kinds. Instead, we might be dealing with socially constructed kinds, wrongly assuming that there are underlying causes and/or essences that explain similarities.

Take race as an example. For a long time, and until rather recently, the received view was that racial differences were securely grounded in genetic or biological differences. Accordingly, different races were thought of as different natural kinds. However, genetic and biological research has subsequently shown that there are no significant genetic differences between various races.

---

[9] For more details see Foster-Hanson and Rhodes (2021).

Differences between races turn out to be based on characteristics that are no more significant biologically than differences in eye or hair color. So why have we assumed that races are natural kinds? Foster-Hanson and Rhodes point to cultural factors being important in many cases of kind-formation. It is telling that in the case of race there is no single generally accepted classification system of races. The U.S.A., for example, lists fourteen race options on some official documents together with "other" in case none of the fourteen applies. Brazil lists five options. South Africa lists four. In each of these cases one can likely, as Foster-Hanson and Rhodes point out, uncover complex cultural and social reasons for the classification being what it is. In the case of race, the *same as* relation is typically grounded in morphological or social similarities. However, what exactly counts as sufficient morphological or social similarities depends, as Foster-Hanson and Rhodes point out, on many factors. As culture and society change, interest in classification changes as well and so a race category might be added, or the criteria of who belongs to a given race might change. When a kind is fluid, so is the *same as* relation.

When Sawyer argues that we need classical style concepts to account for a stable subject matter when discussing, e.g., whales, then that invites a problem. Following that argument, we should have argued a few decades ago that we needed a classical style concept of race to account for what we then viewed as a stable subject matter when discussing races. But, as we now have seen, the concept *race* is in all likelihood a social construct, meaning that we need to account for the subject matter without introducing classical concepts. Similarly, a few decades ago we should have, following her argument, acknowledged that classical style concepts are needed to account for what then was viewed as a stable subject matter when discussing issues related to gender. However, as is now widely acknowledged, the concept *gender* is a social construct and not a concept of a natural kind and so a classical style concept is not called for.

So, if *race* and *gender* are social constructs then what kinds of concepts are they, and what is it that secures sameness of topic when talking about race and gender? Regarding the first question, both prototype style concepts and exemplar style concepts can account for how we group people into races and gender. And regarding the second question, we can still rely on the *same-as* relation to do a lot of the work when it comes to accounting for sameness of topic. [10] When people believed that races were natural kinds, the *same-as* relation was ultimately thought to be grounded in genetic differences. Now when we think that races are social constructs, the *same-as* relation might be grounded in morphological and/or social features that might serve as a prototype or as an exemplar. If the concept *race* is a social construct, then races as natural kinds do not exist. But that does not mean that we cannot group people together based on other features. As social constructs, races do exist and are as real as other social constructs, such as countries and marriages.

While the *same-as* F might go a long way in many cases towards securing sameness of topic, other factors play varying and important roles. One important factor in determining sameness of topic is shared meaning (common ground), i.e., similar representations of the relevant kinds. But in addition to shared meaning, successful communication often requires that the speaker and her respondent make educated assumptions about each other's representations and that they adjust their speech accordingly. For example, if I know that my respondent is only familiar with the smallest kinds of whales, such as the vaquita porpoise and beluga, then I will not start a conversation about whales with her by talking about the giant beasts of the ocean. Instead, I adjust my speech to what I know and/or assume about my respondent's representations.[11] By so doing I make it more likely that we are not talking past each other and that, instead, we are discussing a shared topic. As the

---

[10] Perhaps we can think of the reference-determining properties as homeostatic clusters of properties of the kind that Richard Boyd introduced (1988; 2021).

[11] For more on this see, for example, Geirsson (2013) and Richard (2019).

conversation continues, we might add information about whales that changes our representation of them. For example, my respondent might now include giant whales in her whale concept. Such conceptual change is indicative of increased understanding and knowledge, and it results, if anything, in our respective representations aligning more closely with each other and when doing so making it easier to secure sameness of topic.

4. Drifting and Engineering

The concept *meat* has a history that contrasts in an interesting way with the concepts *gender* and *race*. Until about the 15th century the word 'meat' was used to refer to food, generally. Bread, soup and vegetables all counted as meat. But then the meaning of 'meat' in the English language starts to change and gradually 'meat' only applied to the flesh of warm-blooded animals used for consumption. Nowadays there is also talk about plant-based meat and so the concept is still changing. But while the meaning of 'meat' drifted significantly in the English language, it remained more stable in various Nordic languages. 'Mat' in Swedish and Norwegian and 'matur' in Icelandic and Faroese still means food, generally, and not restricted to animal flesh. While the concept changed in the English language, there was no conscious effort to change the meaning of 'meat'. Instead, the meaning of the term, and the content of the concept, gradually changed over time. Call such unintended development of meaning and content drifting.

While the meaning of 'meat' drifted in the English language, there was no conscious effort to prevent such drift in the Nordic languages. But that is not to say that the meaning of the name and the contents of the concepts remained unchanged. As should be evident now, the contents of many kinds of concepts are fluid and, when it comes to prototypes and exemplars, vary both from one person to the next, as well as over time. How early Icelanders and Faroese represented the concept

matur (food) during the Middle Ages is very different from how they do so now. Food used to be scarce and there was not much variety. Today Icelanders as well as the Faroese have access to food from around the world and the concept *matur* is consequently that much richer. Exemplars, i.e., representations of what falls under the concept, nowadays include everyday food items that the medieval person did not know existed.

When the *same-as* relation is grounded in a natural kind, then the kind is extremely stable. Since the natural kind does not change, it serves as a stable guide to our representations of it. As we discover more about water, our representations of it change. But it is the kind itself that serves as a guide to what representations are accurate and/or not accurate. The situation is very different when it comes to kinds that we mistakenly assume are natural kinds, or kinds that we introduce as social constructs. Here the nature of the kind is very likely not stable. As a result, the members of the kind do not guide our representations to the same extent as natural kinds. Due to the kind being a social construct, the membership in the kind is likely to have soft boundaries and so the membership is likely to change over time. The soft boundaries of many socially constructed kinds allow for changes to prototypes, and they allow for new exemplars to be introduced. The soft boundaries thus allow for a change in membership of the kind, and such change is one element of a meaning change that accounts for the drifting in meaning of certain concepts, such as the concept *meat*.

Drifting is clearly different from engineering. The former involves unintended changes, while the latter involves us consciously trying to change the concept in question. However, how one engineers depends on what kind of concept one is engineering. Let us look at two examples; on the one hand the concept *offside* as used in football (soccer), and on the other hand the concept *woman*.

The concept *offside* is an example of a concept that is socially constructed and defined clearly and precisely. Its meaning is given by a definition, necessary and sufficient conditions. Since the

introduction of the concept, its meaning has been changed multiple times. There is a body of elected officials that decides what counts as offside and that, accordingly, modifies the definition. Over time the changes of the definition of what counts as offside are very significant; so much so that a modern-day footballer would not understand why an offside is called if the call is based on the original offside rules. Once a modified definition is in place, the changes are announced to referees and football fans and the new rules are enforced on the field.

It is interesting, though, that the average football fan recognizes (most instances of) offside when he/she sees them, but is at the same time unable to provide a clear definition of the concept.[12] The people who provide the definition are providing necessary and sufficient conditions, while the average fan is in all likelihood operating with prototype or exemplar concepts, employing those types of concepts when classifying potential offside instances. Still, the change in definition, when enforced on the field, results in the average fan modifying his/her concept of offside.

Someone might point out that here we have a clear example of a concept (although a socially constructed concept), stated with necessary and sufficient conditions, and fans who do not have a full grasp or a full understanding of the concept. So, instead of the members of the body that provides the definition and the fans operating with different kinds of concepts, we only have one type of concept that is well defined but imperfectly understood by the average fan. But, the rejoinder goes, the average fan, when deciding whether someone is offside, does not seem to recall and/or apply a definition to decide the case. Instead, the classification of someone being offside seems based on comparison to previous cases. That is, the average fan does not have to recall or apply clear and precise definitions when he/she classifies an instance as an offside. It is enough that

---

[12] For example, the average fan tends to think that if an opponent (or an appropriate body part) is beyond the last defender when the ball is passed, then an offside should be called. The official rules call for, not the last defender, but the second-to-last defender.

he/she recognizes an instance on the field as a paradigm, or an exemplar, of offsides. The average fan recognizes offside when he/she sees it.

If the above is correct, then we have different types of conceptual engineering at work when it comes to *offside*. On the one hand, the people who change the definition of *offside* work on, and modify, the necessary and sufficient conditions for someone falling under the concept. The fans, on the other hand, do not possess a concept that is given by necessary and sufficient conditions. Instead, the average fan's concept is modified when he/she watches the game and sees the new rules applied, resulting is modified prototypes or new exemplars for the concept.

Compare the offside case with that of the concept *woman*. Clearly, one does not change the concept *woman* in the same way as we change the concept *offside*. There is no official body that provides necessary and sufficient conditions of what counts as a woman. Also, if one were to provide a definition (as, e.g., Haslanger has done), then there is no way to enforce that people change their understanding accordingly. Perhaps a definition can promote discussion of what falls under the concept, but such discussion is likely to take place in small circles of academics. Still, such discussion can become a part of a greater discussion where the traditional understanding of what it is being a woman is challenged. The changes to *woman* have not, and do not come easy, as prevailing and dominant understandings of a concept often resist change. So, what determines how easy or difficult it is to bring about a change, and what are the limits of a change? That often depends on the culture of the times. Some societies have been much more open to embracing a new understanding of *women*, for example, than other societies. A casual observation indicates that societies that strongly resist a new understanding of *women* are often societies where women's' rights have not advanced much and cultural forces, including religious forces, for status quo are strong.

It is important to note that the concept change for *women* does not happen overnight. Instead, the change is a process that takes years, and perhaps decades. During that time the (prototype or exemplar) concepts gradually adjust so that they include representations of individuals that previously would not have counted as proper prototypes or exemplars. We are not dealing with an abrupt change in a definition of a concept, a change in necessary and sufficient conditions, which leads to worries about topic change. Instead, we are talking about gradual modifications of representations, making the change very slow.

An important part of changing the concept *woman* is challenging the traditional understanding in a way that can change prototypes and add exemplars. It is likely that we have over time developed a prototype or an exemplar concept *woman* where individual and cultural experiences determine the content of a concept. While a definition might play a part in that change, encountering people who challenge the dominant prototype or provide new exemplars is more likely to gradually affect a change. Same is likely true of the concept *marriage*. One can try to redefine the concept marriage, but it is likely that the most important factor involved in the concept modification happens when one starts to encounter people whose lifestyles challenge the traditional notion of marriage, thus influencing our prototypes and exemplars.

The cases of *offside*, *woman*, and *marriage* show that conceptual engineering is not a simple and uniform process. In the case of the concept *offside* the concept is engineered top-down; first a body of people changes a definition, and once the new definition is enforced on the playing field, then concepts that the fans have are modified accordingly. In the case of the concepts *woman* and *marriage*, on the other hand, it is likely that social and cultural changes result in the modification of the relevant concepts. The engineering is, in a way, bottom-up. But, just as with the offside example, the engineering efforts for the latter two might be very deliberate.

What the examples show is that conceptual engineering does not always involve working directly on concepts. Instead, conceptual engineering involves a complex relationship between concepts and what they represent. In some cases, conceptual engineering starts with concept modification that affects the world. The various modifications of the concept *offside* provide an example. But the same example also shows how changes in the world affect concepts, for most football fans' concepts of what counts as offside are modified as they watch the game where the new rules are enforced. Similarly, the concepts *marriage* and *woman* are likely modified in light of people providing new paradigms and exemplars in real life. In those cases, the practice comes first, the concept modification follows.

Mark Richard describes what he calls the A-project and the B-project in conceptual engineering. The A-project is typified by Haslanger's revisionary ameliorative accounts of gender and race. The B-project involves, Richard claims, a simpler task "…of getting theorists to agree that most members of the relevant classes—females, those of Hispanic descent—are indeed systematically subordinated on the basis of being members of those classes…" (2020, 360) While both the A-project and the B-project have its place, I am suggesting one more project—call it the P-project, or the peoples' project. The P-project involves activity that is carried out with it in mind to change prototypes and exemplars. It is not a scholarly project carried out by theoreticians. Instead, it is a project that takes place amongst the people. While both the A-project and the B-project might inform activism, the important point is that, for many concepts, the real conceptual engineering is not brought about by scholars and theoreticians.

5. **Concluding remarks**

There is no simple and uniform account of conceptual engineering. There are different types of concepts, and the different types call for different approaches when we try to engineer them. The classical account of concepts, the one that descends from Plato and Frege, has probably received the greatest amount of attention from philosophers who try to redefine, clarify, or ameliorate the concept by providing new or revised necessary and sufficient conditions for the concept's application. Haslanger's amelioration of the concept *woman* serves as an example. But prototype and exemplar concepts do not supply necessary and/or sufficient conditions for concept application. Also, one person's prototype or exemplars might, and in all likelihood does differ from that of other people. When engineering prototype and/or exemplar concepts then that typically involves, not new necessary and/or sufficient conditions, but instead a change to the prototype or different exemplars. Such changes are likely to take place when one encounters what one takes to be a new exemplar and something that changes one's prototype. One can then engineer concepts by providing experiences that affect the concept accordingly.

While one might be tempted to claim that classical concepts apply to natural kinds and prototype and exemplar concepts apply to socially constructed kinds, that is not the case. It is very likely that most peoples' representations, i.e., concept of, e.g., dog is either a prototype or an exemplar type of concept in spite of dogs arguably being a natural kind. And the offside concept as defined by FIFA officials shows that some socially constructed concepts have the earmarks of classical concepts providing necessary and sufficient conditions for application. However, the offside concept that the average football fan acquires is not one that includes necessary and sufficient conditions. Instead, their concepts are likely to be prototype or exemplar concepts; they recognize instances of offsides when they see it and not by applying a definition.

Biggs, Stephen, and Ranpal Dosanjh. "Pervasive Externalism." In *The Routledge Handbook of Linguistic Reference*, edited by Stephen Biggs and Heimir Geirsson, 309-23. New York: Routledge, 2021.

Boyd, Richard. "How to Be a Moral Realist." In *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord, 181-228: Cornell University Press, 1988.

Boyd, Richard. "The Accommodation Theory of Reference." In *The Routledge Handbook of Linguistic Reference*, edited by Stephen Biggs and Heimir Geirsson, 345-58. 2021: Routledge, 2021.

Cappelen, Herman, *Fixing Language: An Essay on Conceptual Engineering.* New York: Oxford University Press, 2018.

Foster-Hanson, E., and M. Rhodes. "The Psychology of Natural Kind Terms." In *The Routledge Handbook of Linguistic Reference*, edited by Stephen Biggs and Heimir Geirsson, 295-308. New York: Routledge, 2021.

Geirsson, Heimir. *Philosophy of Language and Webs of Information.* New York: Routledge, 2013.

Haslanger, Sally, "What Knowledge Is and What It Ought to Be: Feminist Values and Normative Epistemology." *Philosophical Perspectives* 13 (1999): 499-80.

Haslanger, Sally. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?". *Nous* 34, no. 1 (2000): 31-55.

Haslanger, Sally, "What Are We Talking About? The Semantics and Politics of Social Kinds." *Hypatia* 20, no. 4 (2012): 10-26

Haslanger, Sally, "Going on, Not in the Same Way." In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen and David Plunkett, 230-60. Oxford: Oxford University Press, 2020.

Koch, Steffen. "The Externalist Challenge to Conceptual Engineering." *Synthese* 198, no. 1 (2021): 327-48.

Machery, Edouard. *Doing without Concepts.* New York: Oxford University Press, 2009.

Putnam, Hillary. "The Meaning of 'Meaning'." In *Mind, Language and Reality: Philosophical Papers*, 215-71. Cambridge: Cambridge University Press, 1975.

Richard, Mark. *Meaning as Species.* New York: Oxford University Press, 2019.

Richard, Mark. "The A-Project and the B-Project." In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen and David Plunkett, 358-78. Oxford: Oxford University Press, 2020.

Rosch, E., and C. B. Mervis. "Familyresemblance: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7 (1975): 573-605.

Sainsbury, R. M. "Fishy Business." *Analysis*, no. 74 (2014): 3-5.

Sawyer, Sarah. "The Importance of Concepts." *Proceedings of the Aristotelian Society* CXVIII (2018): 127-47.

Sawyer, Sarah, "The Role of Concepts in *Fixing Language*." *Canadian Journal of Philosophy* 50, no. 5 (2020a): 555-65.

Sawyer, Sarah, "Talk and Thought." In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen and David Plunkett, 379-95. New York: Oxford University Press, 2020b.

Sawyer, Sarah, "Truth and Objectivity in Conceptual Engineering." *Inquiry: An Interdisciplinary Journal of Philosophy* 63 (2020c): 1001-1022.

Strawson, P. F. "Carnap's Views on Conceptual Systems Versus Natural Languages in Analytic Philosophy." In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schlipp, 503-18. London: Open Court, 1963.

Thomasson, Amie L. "A Pragmatic Method for Normative Conceptual Work." In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen and David Plunkett, 435-58. Oxford: Oxford University Press, 2020.