Truth and Objectivity in Conceptual Engineering

Sarah Sawyer, University of Sussex
s.a.sawyer@sussex.ac.uk

Forthcoming in a special issue of *Inquiry* on *Externalism and Conceptual Change* edited by Henry Jackman.

1. Conceptual Engineering

Recent examples of conceptual engineering within the philosophical arena include the proposal by Clark and Chalmers (1998) to extend the traditional understanding of belief, the proposal by Haslanger (2000) to rethink our conceptions of race and gender, and the proposal by Scharp (2013) to reconceive the notion of truth. But there is a good sense in which all philosophical theorizing is at root a form of conceptual engineering, and philosophical attempts throughout the ages to capture the nature of knowledge, evidence, causation, explanation, justice, rights, emotion, consciousness, and so on, count equally as examples. Moreover, examples of conceptual engineering can also be found beyond the confines of philosophy. Indeed, much of scientific theorizing falls under the umbrella of conceptual engineering, as do recent proposals in the social arena to overturn the traditional conceptions of, for example, rape, marriage, and women, where such proposals are largely driven by the desire for social justice and equality. Conceptual engineering has a long history and concerns a wide-ranging and diverse array of topics.

Of late, philosophical attention has turned to the nature of conceptual engineering itself. What exactly is conceptual engineering? What unites the diverse array of cases? It will help to distinguish at the outset a broad sense of conceptual engineering from a narrow sense. In the broad sense, conceptual engineering is a form of theorizing that involves a proposed change in linguistic practice. Sometimes this can take the form of a proposal to eliminate the use of a term on the grounds that it is defective in some way, for example by failing to play the explanatory role it was intended to play (e.g. 'phlogiston', 'élan vital'); sometimes it can take the form of a proposal to introduce a new term on the grounds that it is required for explanatory purposes that have not hitherto been recognized (e.g. 'antimatter', 'epistemic entitlement'); and sometimes it can take the form of a proposal to keep a term that is currently in use, but to revise the current use on the grounds that this would constitute some kind of improvement, whether theoretical, practical or normative. Theorizing that involves a proposed change in linguistic practice of any of these kinds—elimination, introduction or revision—is conceptual engineering in the broad sense. But the paradigms of conceptual engineering around which recent debate concerning the nature of conceptual engineering has centred are to be found in that subset of cases that involve the revised use of a term. Each of

the proposals mentioned at the outset of this paper falls into this category, as do ameliorative projects generally. These are instances of conceptual engineering in the narrow sense.

My characterization of conceptual engineering (in both the broad and the narrow sense) as a form of theorizing is intended to be neutral with respect to different metasemantic frameworks that might be adopted in order to explain the phenomenon of conceptual engineering. Thus at this very general level—the level of characterization rather than explanation—I make no mention of concepts so as to remain neutral with respect to the nature of concepts, including the question of whether conceptual engineering should be framed in terms of concepts at all (cf. Cappelen 2018); and I make no mention of linguistic meaning so as to remain neutral with respect to the nature of linguistic meaning, including the question of whether the meaning of a term is determined by current linguistic practice, temporally-extended linguistic practice (cf. Jackman 1999, 2005) or something else. Cappelen's alternative characterization of conceptual engineering as 'the process of assessing and improving our representational devices' (Cappelen, 2018: 3), being non-committal with respect to the nature of the representational devices in play, is also relatively neutral in this regard. However, the characterization nonetheless presupposes that at least some representational devices are the kinds of things that can or should be assessed and improved, and this turns out to be controversial, particularly with respect to the central, revisionary cases (cf. Ball 2019). Moreover, those who advocate understanding concepts in terms of functions appear to reject the claim that the function of a concept is fundamentally representational. If this is right, then the claim that conceptual engineering concerns specifically representational devices is also controversial. (For accounts that appeal to a concept's functions see for example Haslanger 2000, Brigandt 2010, Prinzing 2018, Nado 2019 and Thomasson Forthcoming.) I prefer to think of conceptual engineering as a form of theorizing, and build a case in favour of this characterization—beyond that of metasemantic neutrality—as we proceed.

Focussing on the paradigm cases, I have said that conceptual engineering in the narrow sense is a form of theorizing that involves a proposal to revise the current use of a term on the grounds that this would constitute a theoretical, practical or normative improvement. A proposal to revise the current use of a term can (with a little idealization) be explicitly stated as a revisionary analysis of the term in question. The analysis of a term is revisionary if it specifies conditions that have to be met in order for an object, process or event to fall into the extension of the term, where the specified conditions differ from the conditions that have been traditionally, or are standardly, associated with the term. The

conditions that have to be met in order for an object, process or event to fall into the extension of a term can be thought of as providing the meaning, or intension of the relevant term. Thus, departing from the aforementioned metasemantic neutrality, a proposal to revise the current use of a term can be understood as an attempt to assign the term a revised meaning, or revised intension. (This understanding is inconsistent with temporal externalism, but I leave further discussion of the view until section 4 below.) Assigning a term a revised meaning or revised intension opens up the possibility of a change in the term's extension, since the set of objects, processes or events that satisfy the conditions specified by the revisionary analysis may differ from the set of objects, processes or events that satisfy the conditions traditionally, or standardly associated with the term. Indeed, a revisionary analysis is sometimes proposed precisely in order to eliminate what is perceived to be a discrepancy between the extension of the term as it is currently used and the extension of the term as it ought to be used. To take an example, the traditional extension of the term 'belief' was assumed to be restricted to states that occurred entirely within the confines of the individual thinker, whereas, according to the revisionary analysis, some states that extend beyond the confines of the individual thinker should also be classified as beliefs. Similarly, the traditional extension of the term 'woman' was assumed to include all and only adult human females, whereas, according to the revisionary analysis, some biological males should also be classified as women, and some adult human females should not be so classified. A revisionary analysis thus departs from standard usage and belief.

But a crucial aspect of such cases is that both those who advocate in favour of the proposed revisionary analysis and those who resist are, and typically take themselves to be, talking and disagreeing about a shared topic: belief, gender, race, truth, rape, marriage, women. Thus conceptual engineering narrowly construed essentially involves topic preservation through semantic change. This is what separates such cases from cases involving 'mere semantic drift'. In cases of mere semantic drift, the use of a term changes gradually over time, but the subject matter changes gradually over time in accordance with the gradual change in use. Mere semantic drift is exemplified by vast numbers of terms including 'meat', 'wicked', 'egregious' and 'awesome'. These terms no longer mean what they once did; but nor do they allow us to talk about the same subject matter. A revisionary analysis, in contrast, provides, at the same time as advocating a change in the use of a term, a revisionary understanding of the relevant phenomenon, where the phenomenon in question is identified by use of the term both before and after the variation in its intension and extension.

Since conceptual engineering narrowly construed essentially involves topic preservation through semantic change, a metasemantic framework adequate to explaining the phenomenon of conceptual engineering must be able to answer the question of how this is possible. This is known variously as the continuity problem, the discontinuity problem and Strawson's Challenge, after the objection raised in Strawson (1963) against the account of conceptual explication given in Carnap (1950). According to Strawson, 'typical philosophical problems about the concepts used in non-scientific 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: 506). Strawson's Challenge as it is discussed in the literature on conceptual engineering is a broadening of Strawson's specific concern about the replacement of non-scientific (philosophical) terms with scientific terms. The broader concern is how there can be topic preservation through semantic change.

The pressing nature of the question, and the apparent difficulty of providing a satisfactory answer, in part explains why recent debate surrounding conceptual engineering has centred on these cases (cf. Cappelen 2018, where the problem takes centre stage and which documents the presence of the concern in the literature). The difficulty can be seen in its starkest form if we make the false, but not-unreasonable assumption that the extension of a term delineates a topic about which the term allows us to talk. If extensions and topics are connected (that is, identified) in this way, then a change in extension of the kind that typically follows from the adoption of a revisionary analysis will *ipso facto* bring about a change in topic. The problem, then, is that there is no sense to be made of conceptual engineering in the narrow sense within a metasemantic framework according to which the extension of a term delineates a topic. But the problem runs deeper. Responding to Strawson's challenge requires a metasemantic framework that not only separates extensions from topics in such a way as to allow a change in the former without a change in the latter, but one that separates them in the radical sense of allowing a term to concern the very same topic at two times, t1 and t2, despite there being no overlap at all between the extension of the term at t1 and the extension of the term at t2. This is because examples of conceptual engineering in the narrow sense include cases where the adoption of the revisionary analysis would have the effect of changing the conditions associated with the term from inconsistent to consistent, and (although this doesn't follow of necessity) of changing the extension of the term from empty to non-empty. Indeed, the intention of achieving this result is precisely what motivates certain cases of revisionary analysis. Scharp's revisionary analysis of 'truth' is a case in point (cf.

Scharp 2013).[1] Even if we reject, as I think we should, Scharp's further claim that 'philosophy is, for the most part, the study of inconsistent concepts' (Scharp 2013: 3), such cases provide central examples of conceptual engineering and cannot be set aside as anomalies.

In a series of articles (Sawyer 2018, Forthcoming a, Forthcoming b), I have offered a response to Strawson's challenge by drawing on an externalist metasemantic framework that distinguishes language from thought. The framework countenances two semantic elements—the meaning, or intension of a term, which connects the term to an extension, and the concept expressed by a term, which connects the term to a topic. The meaning of a term is determined by communal linguistic practice, which can change over time resulting in a change in both the intension and the extension of the term. In a certain range of cases, the concept expressed by a term is, in contrast, determined by non-conceptual relations between thinkers and objective properties. In this range of cases, objective properties contribute to the determination of our concepts and constitute the stable topics about which we talk and think. The distinction between language and thought thus provides a response to Strawson's challenge by explaining how topic preservation through semantic change is possible: under certain circumstances, the concept expressed by a term remains constant while the linguistic practice surrounding the use of the term changes to reflect different understandings of what is, and is represented conceptually as, a single topic.

My response depends on the realist assumption that there are objective properties to which we stand in non-conceptual relations and that in certain cases it is such objective properties that both determine our concepts and constitute the topics about which we talk and think. This realist assumption is a familiar and fundamental presupposition of an anti-individualistic understanding of mental representation (cf. Putnam 1973, Burge 1979, 1986). It is, however, more controversial with respect to social kinds and artefact kinds, both of which depend in some way on our mental states and linguistic practices (cf. Searle 1995, Hacking 1999, Epstein 2015) than it is with respect to natural kinds, which it is plausible to think exist independently of our mental states and linguistic practices. Dependence on mental states and linguistic practices might be thought to threaten the kind of objectivity that my account invokes to explain topic preservation through semantic change. The first aim of the

[1] Scharp himself sees this as the replacement of one concept with another, but this is in part because he thinks of concepts as having constitutive principles. As will become clear from the discussion below, I think talk of constitutive principles should be rejected, and talk of the replacement of one concept of truth for another should be reconstrued in terms of the replacement of one theory of truth for another.

current paper is to show that my response to Strawson's Challenge has a wider application than might be thought, applying to terms for natural kinds, philosophical kinds, social kinds and artefact kinds. A significant range of terms in each of these categories express concepts that represent properties that are objective in the relevant sense of exhibiting stability through changes in beliefs and associated linguistic practice. The second aim of the paper is to show that the revisionary analysis of such terms cannot be adequately explained without appeal to the distinction between language and thought.

The structure of the paper is as follows. In section 2, I elaborate the externalist account of concepts in play. In particular, I explain the externalist distinction between concepts and conceptions, and I suggest that it is often conceptions rather than concepts that are engineered in the process of conceptual engineering in the narrow sense. In section 3, I run through a series of examples to support my claim that revisionary analyses of the kind central to projects in conceptual engineering concern topics which are constituted by objective properties in the relevant sense. In section 4, I argue against two rival externalist metasemantic frameworks, each of which attempts to explain topic preservation through semantic change without appeal to the distinction between language and thought. These are given respectively in Cappelen (2018) and Ball (Forthcoming). I argue that, for different reasons, neither provides an adequate account of conceptual engineering in the narrow sense. This is ultimately because of an assumption, implicit in the first and explicit in the second, that the linguistic practice surrounding the use of a term determines the nature of the property represented. I conclude briefly in section 5.

2. Concepts in Conceptual Engineering

Central to an externalist metasemantic framework is a distinction between concepts and conceptions. Concepts are mental representations that are constituents of thoughts. Conceptions, in contrast, are sets of beliefs. Specifically, the set of beliefs a subject associates with a concept is her conception of the subject matter that her concept represents. Since, according to externalism, concept-possession is determined in part by non-conceptual relations to objective properties and is not determined by conceptions, two individuals can possess the very same concept and yet have different associated conceptions.[2] Burge's

[2] Conversely, two individuals can associate the same conception with different concepts. For example, reinterpreting Putnam's Twin Earth thought experiment as having implications for thought rather than language, Oscar and Twin Oscar can be understood as possessing different concepts—the concepts *water* and *twater* respectively—despite the identity of their associated conceptions. See Putnam (1973) for the original Twin Earth thought experiment as applied to the nature of language.

example of Alf, Alf's doctor and the concept *arthritis* provides a clear example (cf. Burge 1979). Alf and his doctor are assumed to possess the very same concept, *arthritis*, despite having different associated conceptions of arthritis. Alf believes (incorrectly) that arthritis can occur in the muscles as well as in the joints, which in part explains why he believes the arthritis in his knee has spread to his thigh; Alf's doctor, in contrast, believes (correctly) that arthritis can only occur in the joints, which is why he goes on to explain to Alf that the pain in his thigh could not be arthritis. Despite the fact that their associated conceptions of arthritis differ, each of their beliefs nonetheless involves the concept *arthritis*: Alf believes that arthritis can occur in the muscles; Alf's doctor believes that arthritis cannot occur in the muscles. And it is because their beliefs involve the same concept that their disagreement counts as a genuine disagreement about arthritis.

The claim that two individuals can possess the same concept and yet have different associated conceptions goes hand in hand with the claim that an individual can possess a concept even though her associated conception is vague or inaccurate. This is the sense in which a subject can be said to have an incomplete, or partial grasp of a concept. Possession of a concept is one thing; full grasp of a concept is another. The extent to which a subject grasps a concept depends in complex ways on two primary factors—her capacity for correct deployment of the concept across a range of contexts, and the truth of her associated conception. But given that concept-possession is determined by non-conceptual relations to objective properties and independently of conceptions, the extent to which a subject's conception of a given subject matter can be wrong is almost limitless. Thus, for example, it is possible to possess the concepts *cat*, *physical object* and *moral obligation*, while nonetheless believing that cats are robots, that physical objects are ideas in the mind of God, and that one's only moral obligation is to oneself. There are no 'core commitments' associated with words that cannot be given up; there are no 'constitutive principles' grasp of which are essential for concept-possession (contra, for example, Peacocke 1992).

Moreover, just as it is possible for an individual to be radically mistaken about a given subject matter, it is, in many cases, possible for the community as a whole to be radically mistaken about a given subject matter. Thus what we might call a 'communal conception', understood as the set of beliefs that constitute the predominant theory of the relevant subject matter in the community—the 'received view', as it were—is also, in many cases, subject to error and open to correction. It is possible, that is, for a given concept to be widely possessed in the community and yet for no-one to grasp that concept fully. (Cf. Burge 1986, but also Burge 2005, where the origins of the distinction between complete and incomplete grasp of a

concept are traced to Frege's work on the concept of number, for which see Frege 1884. A similar point is made in Williamson 2007 and endorsed under the title 'Anti-Creed' in Cappelen 2018: 63). The possibility of the community as a whole being mistaken about the nature of a property or kind reflects a level of objectivity in the property or kind sufficient to ground the possibility of topic preservation through semantic change. Linguistic meanings (and hence communal conceptions) are determined by agreement, but concepts are not.

The distinction between concepts and conceptions provides an alternative framework within which to understand the revisionary analyses central to the paradigms of conceptual engineering. I said in section 1 above that the analysis of a term is revisionary if it specifies conditions that have to be met in order for an object, process or event to fall into the extension of the term, where the specified conditions differ from the conditions that have been traditionally, or are standardly, associated with the term. The difference between the conditions traditionally associated with the term and the conditions the revisionary analysis associates with the term can be understood as encapsulating different conceptions of the same subject matter. This accords with Ball's claim that certain proposals for revisionary analyses are 'best formulated in terms of competing analyses (rather than in terms of competing concepts)' (Ball Forthcoming: 3). In light of this, the proposal to extend the traditional understanding of belief can be understood, in effect, as the proposal of a new theory about the nature of belief. Similarly, the proposal to reconceive the notion of truth can be understood, in effect, as the proposal of a new theory about the nature of truth. Thus the externalist distinction between concepts and conceptions supports my initial characterization of conceptual engineering in the narrow sense as a form of theorizing that involves a revision in the use of a term. It also affords an understanding of the claim, common in the literature, that the conceptual engineer aims to offer a conceptual improvement, since a revisionary analysis constitutes an attempt to provide a more accurate characterization, or theory, of the subject matter represented by the concept. Of course, given the possibility of radically false communal conceptions, there is no guarantee that a revisionary analysis will in fact provide a more accurate characterization of the subject matter represented by the concept, let alone that it will provide the *correct* theory of the relevant subject matter. The possibility of future revision remains forever open. But improvement, measured in terms of truth, is the aim. This contrasts with an anti-truth sentiment that runs through strands of the conceptual engineering literature, particularly evident in discussions of social kinds and in accounts of conceptual engineering that appeal to continuity in a concept's so-called 'functions'. Nado expresses the sentiment when she says: 'Conceptual efficacy is the chief aim of philosophy; truth and

knowledge, by contrast, are secondary aims at best' (Nado 2019: 4). It is also the point at which Ball and I part company. I return to this issue below.

Note that the distinction between concepts and conceptions tells against the prevalent claim that conceptual engineering involves fixing defective concepts, and thus avoids the question of how a concept which has its representational properties essentially could be revised. In section 1, I characterized conceptual engineering in the broad sense as a form of theorizing that involves a proposed change in linguistic practice, whether that be the elimination, introduction or revision of a term. Let us briefly consider each of these in turn. The elimination of a term from the language can sometimes be motivated by recognition of the fact that it fails to denote a property that was originally hypothesized to exist. In so far as a term eliminated on these grounds can be said to express a concept, it can reasonably be thought of as defective in the sense of being empty, but a concept that is defective in this sense is discarded rather than fixed. Terms such as 'phlogiston' and 'élan vital' provide examples of this kind. Conversely, the introduction of a term into the language can sometimes be motivated by the discovery of a hitherto-undiscovered property. Such cases involve the introduction of a new term to express a concept that represents the newly-discovered property, but although this might be thought of as fixing a conceptual system which is defective as a whole—defective on the grounds that it did not contain a concept capable of representing the property in question—such cases do not involve defective concepts, and no concept is fixed in the process. 'Anti-matter' and 'epistemic entitlement' provide recent examples of this kind. Finally, a revisionary case is often best understood neither in terms of a defective concept that is eliminated nor in terms of a defective conceptual system that is enhanced. Rather, the defect in a revisionary case often lies squarely in the associated communal conception. Talk of inconsistent concepts in the revisionary context, then (cf. Eklund 2002, Spicer 2008, Scharp 2013), should at least sometimes be reconstrued in terms of inconsistent conceptions. It is sometimes the belief that the received theory is incorrect that leads to the proposed revisionary analysis, one that involves the very same concept but encapsulates a different associated conception. 'Belief', 'woman', 'truth' and so on provide examples of this kind.

Moreover, once we reject the claim that conceptual engineering involves fixing defective concepts, there is no need to invoke a concept's functions, purposes or aims in order to secure continuity through semantic change in the general case, although it may still be appropriate in certain cases. Appeal to some aspect of a concept's function has been a popular strategy for responding to the concern that underlies Strawson's challenge. Thus

Haslanger maintains that semantic change is justified 'if central functions of the term remain the same, e.g., if it helps to organize or explain a core set of phenomena that the ordinary terms are used to identify or describe' (Haslanger 2000: 35); Brigandt maintains that the epistemic goal of a concept, understood in terms of 'the kinds of inferences and explanations that the concept is intended to support' (Brigandt 2010: 24) explains the rationality of the semantic change surrounding biological terms such as 'gene'; Prinzing (2018) claims that topic preservation can be secured by appeal to the preservation of a concept's function, where this is marked out by preservation of the concept's 'essential features'; and Thomasson says that 'appealing to function provides a promising way of giving a sense in which we remain on topic across change in intension and extension' (Thomasson Forthcoming: 7).[3] But while appeal to a concept's function may in some cases be illuminating, appeal to a concept's function cannot explain topic preservation through semantic change across the board. This is because a concept's capacity to organize or explain a core set of phenomena, or to fulfil its epistemic goal, often depends primarily on whether it represents an objective property that is instantiated by the relevant core examples. Moreover, the only essential feature of a concept is often its representational content. I leave detailed analysis of such accounts to another occasion, but I share Cappelen's pessimism about their prospects in the general case (cf. Cappelen 2018, chpt. 16). For now, it is sufficient to note that the distinction I have drawn between the linguistic meaning of a term and the concept it expresses, and the corollary distinction between concepts and conceptions, provides an explanation of topic preservation through semantic change without appeal to a concept's function.

3. The Objectivity of Topics in Revisionary Analyses

Putnam's Twin Earth thought experiment (Putnam 1973), in which physical doppelgängers Oscar and Twin Oscar are related to superficially identical but fundamentally different natural kinds—water and twin water respectively—has played a significant role in the current widespread acceptance of externalist theories of language, and, following the developments in Burge (1979), in the current widespread acceptance of externalist theories of thought. The Twin Earth thought experiment presupposes the existence of objective natural kinds to which we are causally related and about which our theories, and hence associated conceptions, may

[3] Nado (2019) also endorses a functional account of concepts, which she calls 'radical functionalism'. According to Nado, continuity of function is all-important, and conceptual engineering need not involve topic preservation at all. This is consistent with what I have said given the distinction I have drawn between conceptual engineering in the broad sense, examples of which Nado uses to support her claim, and conceptual engineering in the narrow sense, which essentially involves topic preservation.

be incomplete, or even fundamentally mistaken. The thought experiment seeks to establish the fact that it is, in such cases, causal relations rather than associated conceptions that determine our concepts. Kripke's discussion of tigers and gold highlights what is essentially the same, anti-descriptivist point, namely that what natural kind terms allow us to talk and think about is not determined by our theories, which may be fundamentally mistaken, but by our causal connections to instances of the kinds in question (Kripke 1972). Thus Kripke describes a scenario in which the characterization of a tiger as ' ... "a large carnivorous quadrupedal feline, tawny yellow in colour with blackish transverse stripes and white belly," (derived from the entry under 'tiger' in the *Shorter Oxford English Dictionary*)' (Kripke, 1972: 119) is radically misguided, being based on observations of tigers in non-optimal conditions which masked their true nature as shy, three-legged herbivores. Discovery of the empirical error would lead to a revised characterization, a revised dictionary entry and a revised communal conception; but, crucially, the concept would remain constant throughout the revisions, as would the topic—tigers. Revisions to theories of such natural kinds thus exhibit topic preservation through semantic change.[4] Such natural kind terms are, as a result, subject to revisionary analysis in the sense described above.

The role of demonstratively-given examples is fundamental to such cases. Kripke says, 'the original concept ... is: *that kind of thing*, where the kind can be identified by paradigmatic instances' (Kripke 1972: 122, original emphasis). The underlying point does not depend on Kripke's portrayal of the concept as itself demonstrative, which we should not take too seriously; similarly for Putnam's claim that 'words like "water" have an unnoticed indexical component' (Putnam 1973: 710). The underlying point is that it is the direct connection to paradigm examples, unmediated by associated beliefs, that allows the concept to represent the stable objective property that it does. The importance of demonstratively-given examples to thought and language is also emphasised throughout Burge's work. For example, in his (1977) he argues for the fundamental nature of *de re* belief to language, thought and knowledge; and in his (1986) he emphasises the role of examples, given both perceptually and imaginatively, to the determination of linguistic meaning.

This is not to say, of course, that the range of examples taken to instantiate a property cannot itself be mistaken; on the contrary, the range of examples taken to instantiate a given natural kind property is itself subject to revision. However, revised judgements about

[4] Kripke does not distinguish the meaning of a term from the concept expressed by the term, but talks indiscriminately of both. As should be clear, I agree with the fundamental point as related to concepts but take the change in theory to mark a change in meaning.

particular examples are significantly constrained in a way that revised associated conceptions are not. Thus a piece of fool's gold might initially have been counted as a piece of gold, but if the paradigm examples had included no gold, or had not been predominantly examples of gold, it is not clear that the concept would have been a concept of gold at all. It is the causal constraint on concept-possession that places the constraint on the extent of possible revision with respect to paradigm examples. It is also worth noting, however, that despite the fundamental role of demonstratively-given examples in concept-determination, the use of a natural kind term in empirical explanations and predictions is typically governed by surrounding theoretical commitments. As a result, the efficacy of empirical explanations and predictions depends not only on having concepts that track natural kinds, but in large measure also on the accuracy of the associated conceptions. It is no surprise that predictions of hurricanes have improved since we have come to know more about the causes and nature of hurricanes—but it is hurricanes we were thinking about all along.

The considerations above apply equally to a range of philosophical terms, such as 'colour', 'number', 'emotion', 'consciousness', 'person', 'physical object', 'validity' and 'truth'. There are concrete examples of each of these, and it is through causal relations to such concrete examples that we acquire concepts that represent the objective properties they instantiate. But causal relations to concrete examples do not reveal the true nature of the properties instantiated. As such, our theories of such properties may be fundamentally mistaken; disagreements about whether colours are mind-dependent, whether numbers are sets, whether emotions are beliefs, and so on, are, respectively, disagreements about the nature of colours, numbers and emotions. The entrenched nature of philosophical disagreement in certain cases highlights the fact that competing characterizations of kinds or properties are possible in the absence of a settled linguistic meaning. A revisionary analysis need not, then, be a rival to a traditional, or standard view, but can be one amongst several competing characterizations. Nonetheless, a revisionary analysis will constitute an improvement if it better captures the nature of the relevant property. Moreover, the very fact that the disagreements are disagreements over the nature of the properties concerned is sufficient to establish that concepts are not determined by associated conceptions, but by causal relations to objective properties. Such concepts can be widely possessed in the community but not fully grasped by anyone.

The considerations surrounding natural kind terms apply even more directly to a wide range of social kind terms. Social kinds are, it is reasonable to assume, dependent in some sense on our linguistic practices and mental states. They are, one might say, socially

constructed, relying for their existence on contingent facts about our social relations. But they are not, for all that, stipulated into existence, and their nature is neither transparent to us nor determined by us. On the contrary, they are objective kinds, open to empirical investigation in much the same way that natural kinds are. The purpose of the social sciences is arguably to study such kinds, and to offer explanatory and predictive theories about them. Social class, for example, is a complex social phenomenon, thought to be determined by a number of socioeconomic factors, primarily income, wealth, education and occupation, and thought to have a significant impact on things such as physical health, life expectancy, and the prospects of one's children. The nature of social class, and its causes and effects, are not stipulated but discovered. Similarly, it is a significant empirical discovery that race is not, as once thought, a biological category, and that it affects, in either a positive or a negative way, one's capacity to enter the professional workforce, the probability that one will suffer life-long systematic discrimination, and one's physical and mental health. The same can be said of warfare, poverty, crime, punishment, gender, and so on: their natures are not stipulated but discovered. Theories of social kinds are thus subject to error and to subsequent revision, where such revisions constitute improvements in so far as they better capture the nature of the kinds in question. But here too it is possible for the community as a whole to be mistaken about the nature of social kinds. This reflects a level of objectivity sufficient to ground the distinction between social kind concepts and our conceptions of the social kinds they represent, and it establishes that the former are not determined by the latter.[5] Similar considerations apply to terms for artefacts, which also depend in some sense on our mental states, linguistic practices and social interactions. (Cf. Burge 1986, in which the distinction between thought and linguistic practice is illustrated with reference to the term 'sofa'. See also Sawyer Forthcoming b.)

In each of these categories, some terms are further removed from paradigm demonstratively-given examples than others. Natural kind terms such as 'quark', 'gravity' and 'gene', philosophical terms such as 'concept', 'haecceity' and 'supervenience', and social terms such as 'collective intentionality', 'ideology' and 'modernity' provide examples. It might be tempting to focus on the role of surrounding theoretical beliefs in the introduction of such terms, and hence treat such terms as in some sense 'purely descriptive', and hence either not subject to revisionary analysis, or subject to revisionary analysis understood along

[5] Social terms also often have a moral dimension. If some form of moral realism is true, which I think it is, we have reason to revise our understanding of social kinds so that we do not act in contravention of the moral facts.

different lines. However, the connection to what is demonstratively given provides an anchor to the observable world, lending some plausibility to the claim that even revisions to theories of unobservable kinds can sometimes exhibit topic preservation through semantic change of the kind I have been advocating. Thus it is plausible to think of the discovery that neutrinos oscillate between three different 'flavours' as they travel—electron, muon and tau—and are thus not massless, as once thought, but instead have an immeasurably tiny mass, as a discovery about neutrinos. This requires thinking of the discovery as leading to a revised communal conception associated with a stable concept capable of securing topic preservation through semantic change. Two points favour the realist interpretation. First, although the possibility of reference-failure may increase as our terms are more theoretically-infected because further removed from demonstratively-given paradigm examples, the possibility of reference-failure is nonetheless present even at the most basic level. Just as it turned out that there was no phlogiston, it may have turned out that there were no tigers, not in the (illegitimate) sense that there were no creatures satisfying the specified conditions, but in the sense that there were no creatures responsible for our observations. The possibility of reference-failure, then, does not in and of itself undermine the objectivity of the properties referred to by the terms in a given category. Second, the descriptivist interpretation arguably involves a commitment to scientific anti-realism, which, given its descriptivist roots, looks to be inconsistent with the realist assumptions underlying the externalist view.[6] There is no principled reason, then, to think that the kinds of terms that exhibit topic preservation through semantic change need be restricted to terms that have demonstratively-given examples, so long as the kinds of cases that involve demonstratively-given examples are seen as primary and fundamental to higher-level theoretical thought.[7]

It is common in the literature to distinguish conceptual engineering from other forms of theorizing, contrasting it both with empirical, scientific theorizing and with the traditional philosophical endeavour of conceptual analysis. Given my understanding of conceptual engineering, it should be clear that I reject the former contrast. I also reject the latter. Nado captures the widely-assumed contrast when she says: 'Conceptual engineers aim to improve

[6] Externalism is clearly inconsistent with the scientific anti-realism of the logical positivist era, which is descriptivist through and through. Although I do not have the space to develop the argument here, there is reason to think that externalism is also inconsistent with the kind of scientific anti-realism found in van Fraassen (1980, 1989).

[7] Sometimes we improve our incomplete understanding of a messy world by creating precise models which we can understand completely but which do not represent the world accurately. I do not have the space to explore this issue here, but note that the phenomenon occurs both in philosophy and in science. For example, with respect to philosophy this is one way to understand Carnap (1950) on explication, and for the claim that scientific laws do not represent the regularities that occur in the physical world see Cartwright (1983).

or replace rather than to analyse; to create rather than to discover. While conceptual analysts are interested in the concepts we *do* have, conceptual engineers are interested in the concepts we *ought* to have. The project is prescriptive rather than descriptive.' (Nado 2019: 3, original emphasis). If I am right, however, there is no substantial difference between conceptual analysis and conceptual engineering. Conceptual analysis was never literally an analysis of concepts—it was always an attempt to understand the nature of, for example, knowledge, evidence, causation, explanation, justice, rights, emotion, consciousness, truth, even if it was not understood as such. Similarly, conceptual engineering is not literally the engineering of concepts—it is equally an attempt to understand the nature of knowledge, evidence, causation, explanation, justice, rights, emotion, consciousness, truth, even if it is not generally understood as such. Moreover, conceptual analysis and conceptual engineering are alike interested in the concepts we do have and the concepts we ought to have—in improving and replacing as well as in analysing. The distinction between conceptual engineering in the broad sense and conceptual engineering in the narrow sense helps to make this clear.

In a more speculative spirit, the label 'conceptual analysis' may have been used, at one time, with the intention of distinguishing philosophy from science by appeal to an alleged distinction between conceptual matters and empirical matters, where the former were to be investigated by an appeal to meaning, and the latter were to be investigated by appeal to empirical observation. But the alleged distinctions were called into question by Quine's insight that there is no separating truths of meaning from matters of fact (cf. Quine 1951). Quine's insight informs Burge's account of meaning, according to which a statement of meaning is inextricably connected to what the community takes to be a statement of the facts (cf. Burge 1986). It also informs my account of revisionary analysis, according to which it involves both a change in the use, and hence meaning, of a term, and, at the same time, offers a revisionary understanding of the relevant subject matter. But it is the further realization that we can err in what we take the facts to be that leads Burge to draw the crucial distinction between thought and language that I endorse (cf. Burge 1986). When we state what we take the facts to be, we are responsible to the nature of the facts themselves. We can go wrong not just individually but collectively. This is the sense in which we are guided by norms that go beyond actual linguistic practice, and hence by norms that go beyond not just individual but communal conceptions.

4. Topics and Linguistic Practice

I have argued that conceptual engineering in the narrow sense is to be explained by appeal to the externalist distinction between concepts and conceptions. In particular, if a concept is determined by non-conceptual relations to an objective property, rather than by associated communal conceptions, topic preservation through semantic change will be possible. The level of objectivity required for topic preservation through semantic change is guaranteed by the mere possibility of collective error; it does not depend on a stronger level of objectivity, such as mind-independence or independence from linguistic or social practice more generally. Moreover, the requisite level of objectivity is, I have argued, exhibited not only by natural kinds, but also by a wide range of philosophical kinds, social kinds and artefactual kinds. In this section, I argue against two rival externalist metasemantic frameworks, each of which offers an alternative explanation of conceptual engineering in the narrow sense. The first is the 'worldly construal' of conceptual engineering given in Cappelen (2018); the second is the account of conceptual engineering offered in Ball (Forthcoming), inspired by temporal externalism as proposed and defended in Jackman (1999, 2005). I briefly look at each in turn.

Cappelen endorses what he calls 'the worldliness of conceptual engineering', according to which the process of conceptual engineering changes not only the intensions and extensions of our terms, but also the world. He says: 'On this view, an instance of successful conceptual engineering, e.g. of 'person', has the result that what a person is has changed' (Cappelen 2018:46). I have argued elsewhere that Cappelen's account of conceptual engineering fails to explain the possibility of topic preservation through semantic change (cf. Sawyer Forthcoming b). Since, according to Cappelen, conceptual engineering changes what a person is by changing the conditions that have to be satisfied in order for an object to fall into the extension of the term 'person', the property of being a person is not a stable objective property that exists independently of the conditions associated with the term 'person'. The worldly construal of conceptual engineering thereby undermines topic preservation by eradicating the possibility of a single topic—persons, for example—that persists throughout the change. Note that Cappelen does not confine his claims to social terms, which some might find plausible, but goes on to say: 'I think this 'worldly' description is the correct way to describe all instances of conceptual engineering, not just in the social domain' (Cappelen 2018: 46), mentioning specifically its application to natural kind terms. This means that, for example, if we were to discover, as in Kripke's hypothetical scenario, that tigers are in fact shy, three-legged herbivores, and we adopt a revisionary analysis of the term 'tiger' to reflect this discovery, we would have changed the *nature* of tigers. This implication is highly

counter-intuitive. But more importantly for present purposes, we are now in a position to see how antithetical to externalism the view is.

To think that a change in the conditions associated with the term 'tiger' could bring about a change in the nature of tigers, is to embrace a descriptivist theory of reference according to which what we talk and think about depends on descriptions we have in mind. It is precisely this kind of descriptivist account of reference that the externalist theories of Putnam, Kripke and Burge were intended to overturn. Note that Kripke's claim is not merely that the description an individual associates with the term 'tiger' might be false, but that the description the community as a whole associates with the term 'tiger' might be false. This is evident from his appeal to the dictionary definition of 'tiger'. The externalist insight is that our capacity to think and talk about objective properties—indeed, our capacity for representation *per se*—depends on our non-conceptual relations to objective properties about which we may be collectively mistaken. The claim that the process of conceptual engineering changes the nature of tigers reverses the direction of the determination-relation that the externalist advocates; it makes the nature of tigers dependent on our conceptions rather than our concepts dependent on the nature of tigers. But tigers are what they are, and are not made so by any agreement we may collectively reach. The descriptivist approach undermines the objectivity of the properties referred to by our terms.

The same descriptivist reversal of the determination-relation is an unwelcome implication of the account of conceptual engineering advocated by Ball (Forthcoming). Ball maintains, first, that we should not treat revisionary analyses as if they change the subject. Subject-change views, he says 'cannot explain our argumentative practice: [they] cannot make sense of the kinds of arguments we offer, and the way we respond to these arguments' (Ball, Forthcoming: 4). I agree, and this claim accords with my characterization of revisionary analyses as involving topic preservation through semantic change. However, Ball also maintains, second, that revisionary analyses do not involve semantic change. Drawing on work on temporal externalism (cf. Jackman 1999, 2005), Ball claims that 'our theoretical activity shapes what we mean, but it does so not by making us mean something new, but by shaping what we meant all along' (Ball, Forthcoming: 3). A successful revisionary analysis—successful in the sense of being accepted by the community after debate and reflection—is, on this view, a retrospective stipulation of meaning. He says: 'the key to understanding revisionary analyses is that they involve stipulation that (partially) fixes the meaning of prior uses of the word: in this kind of case, *first* we go about using a word, then *later* we make the stipulation that gives meaning' (Ball, Forthcoming: 13-14, original emphasis). Thus the

liberal way in which we now use the term 'marriage' in part determines the meaning that the term 'marriage' had all along. Similarly, whether Haslanger's revisionary analysis of the term 'woman' (cf. Haslanger 2012) correctly captures what we currently mean by the term 'woman' depends, ultimately, on whether her analysis is accepted, on reflection, in the future. This second, temporal externalist claim is to be accepted in preference to what Ball sees as the only alternative once we reject the subject-change view. That is, rejecting the subject-change view leaves us, according to Ball, with only two options. Either the facts that fix meaning must be in place at the beginning of a discourse, or the facts that fix meaning can be determined by the discourse itself. This is why he says: 'it looks like we have a choice of preferring our views *prior to inquiry* or our views *after inquiry*: but surely (other things equal) our views after inquiry are to be preferred' (Ball, Forthcoming: 16, original emphasis).

Ball is, of course, right that our views after inquiry are to be preferred. This is because our views after inquiry are more likely to reflect the true nature of the properties about which we think and talk. But this does not mean that our views after inquiry should be taken to *determine* the topics about which we think and talk. As with Cappelen's view, this is to fall back into a descriptivist account of reference. Stipulations made on the basis of agreement after rational reflection will always in principle be open to challenge, no matter how much evidence has been gathered in their favour. Why, after all, couldn't future members of our linguistic community be collectively mistaken? Some properties and kinds are what they are, and are not made so by any agreement we may collectively reach, whether now or in the future, even after we have reflected for an indefinite amount of time on all the evidence available to us. The possibility of communal error remains, but it is ruled out by fiat on both Cappelen's view and on Ball's.

5. Conclusion

I have argued that conceptual engineering in the narrow sense is to be explained by appeal to the externalist distinction between concepts and conceptions. If a concept is determined by non-conceptual relations to an objective property, rather than by associated communal conceptions, topic preservation through semantic change will be possible. The level of objectivity required for topic preservation through semantic change is guaranteed by the mere possibility of collective error and does not depend on a stronger level of objectivity, such as mind-independence or independence from linguistic or social practice more generally. This means that the requisite level of objectivity is exhibited not only by natural kinds, but also by a wide range of philosophical kinds, social kinds and artefactual kinds. Alternative accounts

that fail to distinguish language from thought thereby fail to do justice to this basic level of objectivity, and subsequently fail adequately to explain the phenomenon of conceptual engineering in the narrow sense.

Acknowledgements

With thanks to Joachim Horvath, Steffen Koch, Guido R. Löhr and Albert Newen for helpful discussion following presentation of this paper at the EXTRA Research Colloquium at Ruhr-Universität Bochum.

References

Ball, D. 2019. "Fixing Language: An Essay on Conceptual Engineering, by Herman Cappelen." *Mind* https://doi.org/10.1093/mind/fzz011

Ball, D. Forthcoming. "Revisionary Analysis without Meaning Change (Or, Could Women be Analytically Oppressed?)." In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen and D. Plunkett. Oxford: Oxford University Press.

Brigandt, I. 2010. "The Epistemic Goal of a Concept: Accounting for the Rationality of Semantic Change and Variation." *Synthese* 177 (1): 1211-1241.

Burge, T. 1977. "Belief *De Re*." *Journal of Philosophy* 74 (6): 338-62.

Burge, T. 1979. "Individualism and the Mental." In *Midwest Studies in Philosophy 4*, edited by P. French, T. Uehling, and H. Wettstein. Minnesota: Minnesota University Press: 73-122.

Burge, T. 1986. "Intellectual Norms and Foundations of Mind." *Journal of Philosophy* 83 (12): 697-720.

Burge, T. 2005. *Truth, Thought, Reason: Essays on Frege.* Oxford: Oxford University Press.

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

Carnap, R. 1950. *The Logical Foundations of Probability*. University of Chicago Press.

Cartwright, N. 1983. *How the Laws of Physics Lie.* Oxford: Oxford University Press.

Clark, A. and Chalmers, D. 1998. "The Extended Mind." *Analysis* 58 (1) 7-19.

Eklund, M. 2002. "Inconsistent Languages." *Philosophy and Phenomenological Research* 64 (2): 251-275.

Epstein, B. 2015. *The Ant Trap: Rebuilding the Foundations of the Social Sciences.* Oxford: Oxford University Press.

Frege, Gottlob. 1884. *Die Grundlagen der Arithmetik: Eine logisch-mathematische Untersuching uber den Begriff der Zahl.* Breslau: Verlage Wilhelm Koebner. Translated by J. L. Austin as *The Foundations of Arithmetic*: *a logico-mathematical enquiry into the concept of number.* Oxford: Basil Blackwell, 1950.

Hacking, I. 1999. *The Social Construction of What?* Cambridge, Mass.: Harvard University Press.

Haslanger, S. 2000. "Gender and Race: (What) are they? (What) do we want them to be?" *Noûs* 34 (1): 31-55.

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

Jackman, H. 1999. "We Live Forwards but Understand Backwards: Linguistic Practices and Future Behaviour." *Pacific Philosophical Quarterly* 80: 157-177.

Jackman, H. 2005. "Temporal Externalism, Deference, and our Ordinary Linguistic Practice." *Pacific Philosophical Quarterly* 86: 365-380.

Kripke, S. 1972. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.

Nado, J. 2019. "Conceptual Engineering, Truth, and Efficacy." *Synthese* https://doi.org/10.1007/s11229-019-02096-x

Peacocke, C. 1992. *A Study of Concepts.* Cambridge, Mass.: MIT Press.

Prinzing, M. 2018. "The Revisionist's Rubric: Conceptual Engineering and the Discontinuity Problem." *Inquiry* 61 (8): 854-880.

Putnam, H. 1973. "Meaning and Reference." *The Journal of Philosophy* 70: 699–711.

Quine, W.V.O. 1951. 'Two Dogmas of Empiricism'. Philosophical Review, 60, pp. 20–43.

Sawyer, S. 2018. "The Importance of Concepts." *Proceedings of the Aristotelian Society* 118 (2): 127-147.

Sawyer, S. Forthcoming a. "Talk and Thought." In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen and D. Plunkett. Oxford: Oxford University Press.

Sawyer, S. Forthcoming b. "The Role of Concepts in Fixing Language." *Canadian Journal of Philosophy.*

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

Searle, J. 1995. *The Construction of Social Reality*. Free Press.

Spicer, F. 2008. "Are There Any Conceptual Truths about Knowledge?" *Proceedings of the Aristotelian Society* 108 (1): 43-60.

Strawson, P. F. 1963. “Carnap’s Views on Conceptual Systems versus Natural Languages in Analytic Philosophy.” In *The Philosophy of Rudolf Carnap*, edited by P. A. Schilpp. Open Court: 503-518.

Thomasson, A. Forthcoming. “A Pragmatic Method for Conceptual Ethics.” In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen and D. Plunkett. Oxford: Oxford University Press.

van Fraassen, B. 1980. *The Scientific Image.* Oxford: Clarendon Press.

van Fraassen, B. 1989. *Laws and Symmetry*. Oxford: Clarendon Press.

Williamson. T. 2007. *The Philosophy of Philosophy.* Oxford: Blackwell.