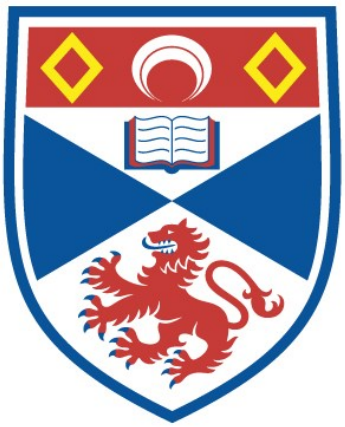
Reality Engineering and social kinds

Duccio Calosi

A thesis submitted for the degree of MPhil
at the
University of St Andrews



2022

Candidate's declaration

I, Duccio Calosi, do hereby certify that this thesis, submitted for the degree of MPhil, which is approximately 39,990 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2019.

I confirm that no funding was received for this work.

Date 23 Sept 2021 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of MPhil in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date 22 Sept 2021 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Duccio Calosi, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date 23 Sept 2021 Signature of candidate

Date 22 Sept 2021 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Duccio Calosi, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date 23 Sept 2021 Signature of candidate

Contents

Reality Engineering and Social Kinds

Abstract

Conceptual Engineering is a new and interesting trend in Philosophy. However, it is not free from problems. The most relevant issue is that, at least following a Cappelen-like account, we are forced to commit to the controversial metaphysical view that the world has a linguistic structure. Under such view, a modification in the semantics of a term implies a modification in nature of the thing which is referred by that word. I propose to explore the implications of the reversal of such principle, thereby committing to the idea that a modification in the nature of things implies a modification of the semantics of the terms that refer to them, and not the other way around. Following this new principle, I am interested in developing an alternative account to Conceptual Engineering, which I call (following Greenough) Reality Engineering.

In this dissertation, I will focus on the analysis of two major points about Reality Engineering, trying to define what it is about and how to perform it. I will argue that Reality Engineering has kinds as its scope and I will restrict the focus of the present enquiry to social kinds only. I will proceed by providing a taxonomy of the most popular views about the metaphysics of social kinds, since in order to modify something properly, first we have to be clear on what that something is. Out of this taxonomy, I will generate two general theories on social kinds. The first one is what we can call a Top-down view, and it says that a social kind is generated via the acceptance of constitutive rules by some group of authorities and the successful application of those rules in ordinary practice. The second one is what we can call a Bottom-up view, according to which social kinds are nothing but the reification of social external norms,

where social external norms are to be intended as the set of attitudes/behaviours/treatments/practices that people who are not members of the kind have towards the members of such kind (trivially, if the kind in question, like *money*, does not include people as its members, then everyone is external to such kind).

After presenting these two views, I will explore the possibility of engineering kinds within them, focusing on some case studies and examples. I will highlight various ways in which social kinds can be defective and propose solutions for all kinds of defectiveness.

In conclusion, I will briefly discuss how typical worries concerning Conceptual Engineering projects translate to my framework, focusing on the problems of Feasibility and Control.

Chapter 1 – Conceptual Engineering: What is it and Why to care about it

Conceptual engineering is a philosophical discipline that aims to ameliorate our representational devices. This movement is motivated by the observation that some of our concepts prove to be defective, and the will to fix this defectiveness. There are multiple ways in which a concept can be defective, as Patrick Greenough points out in his paper from 2019 (Greenough, 2019, p. 408). They might be inconsistent, in the sense that they might have contradictory conceptual principles. Work on this kind of defectiveness can be found in Scharp's engineering project about *truth* (Scharp, 2013). Concepts could also be not useful for the phenomena that they wish to track, like in the case of Clark and Chalmer's treatment of *belief* (Clark and Chalmers, 1998). Also, concepts might generate problematic social or moral effects, like we can see in Haslanger's account of *gender* (Haslanger 2005, 2006). This are just few examples, but they are useful to appreciate the fact that Conceptual Engineering focuses on many areas, within and outside philosophy, ranging from theoretical philosophy (Scharp) to the philosophy of mathematics (Tanswell, 2018), social philosophy (Haslanger) and more.

Conceptual Engineering is also a metaphilosophical guideline, that wishes to reorient philosophy towards new normative work, aiming for more practical and active engagement with the various sub-disciplines. In the literature, much more work has been done on the development of the metaphilosophical side of the discipline rather than on individual projects. Notable examples are Scharp's account of philosophy as the study of defective concepts (Scharp, 2020), Cappelen's discussion of the general goals

and methods of the discipline (Cappelen, 2018), Belleri's analysis of the role of Conceptual Engineering with respect to ontological disputes (Belleri, 2020), Thomasson's account of *functions* of concepts as the matter of Conceptual Engineering (Thomasson, 2020), and Plunkett's account of Conceptual Ethics (Plunkett, 2015), and more.

As we can already grasp, Conceptual Engineering is a rather broad movement, both in terms of scope and in terms of methods. However, we can still break down a general pattern that unifies most of the view about the stages of engineering projects and their goals. This is relevant because it will help us having a better grasp of the nature of the discipline and appreciating the reasons why we should take these matters seriously. Let me then bullet the three main stages of such pattern:

- The first stage of engineering projects is always ***evaluation***. In this stage, we look for defectiveness in a concept, which can be present in different ways. Also the scope of evaluation may vary. For instance, Cappelen's idea is that we should look for defectiveness in meanings of words (Cappelen, 2018), Scharp looks for defectiveness in conceptual principles (Scharp, 2013) and Thomasson analyses how concepts behave with respect to the function that we want them to serve (Thomasson, 2020). The common point is that our representational devices are defective, and the evaluation stage should locate and explain such defectiveness.
- Then comes the stage of ***fixing*** this defectiveness. Also in this case we might take different paths. The most important distinction in the literature is between *revisionary* projects and *replacement* projects. Revisionary projects aim to keep the old concept by getting rid of its defectiveness and maintaining its good

features. Replacement projects aim to substitute the old defective concept with a new unproblematic one. A further option is represented by *elimination* projects, which aim to get rid of a problematic concept once and for all, without even substituting it. Also, we might have *introduction* projects, that aim to create a new concept that might fill some gaps in our representational apparatus. All of these options are respectable and interesting but taking a stance on which one is better is not a purpose of the present research.

- The last stage of engineering projects which is common to most accounts of the discipline is the ***implementation*** stage. This is the stage in which the new concept is applied and starts to be used by the relevant communities. There is no doubt that this stage has proven to be the most difficult to deal with for Conceptual Engineers, since there is almost no instance of engineering projects (within philosophy at least) that have obtained the desired relevance. I am not in a position to analyse the problem too deeply, but I will provide an account of the reasons why this issue emerges at the end of the next chapter.

At this point, we are in a position to appreciate the relevance of Conceptual Engineering as a movement. The promises of Conceptual Engineering are extremely appealing, since it seems that we have a tool to improve both our theories and our practices. The main point of the discipline is that we should not be happy with defective concepts, we should find a way to ameliorate them. Within and outside of philosophy, this is without a doubt a noble enterprise.

It is doubtful, however, that Conceptual Engineering can deliver on its promises. This dissertation is, indeed, just about this. I will argue that Conceptual Engineering is not able, by itself, to accomplish its goals, and I will look for an alternative account.

Chapter 2 – Conceptual Engineering and the World

2.1 Cappelen's Conceptual Engineering

In order to analyse the relation between Conceptual Engineering and the world, I will start off by analysing Herman Cappelen's view on the topic. Cappelen has emerged as one of the most prominent philosophers in the Conceptual Engineering movement. His view on the subject, however, is quite unique, since he starts off by contesting that Conceptual Engineering should focus on concepts at all (Cappelen, 2018).

Cappelen argues, in a rather promising way, that Conceptual Engineering is better off focusing on **changing the meaning of terms instead of concepts** (see Cappelen, 2018, chapter 5)**.** This idea follows from his goal of formulating a general theory for Conceptual Engineering, without too much focus on individual examples. On such a general standpoint, it is easy to see that focusing on meanings, in terms of intensions and extensions, has clear advantages over focusing on concepts.

The nature of concepts is a topic of deep debate in the philosophical literature, and one could not generate a theory of conceptual change if not by addressing such endless discussions. Also, it is still not clear whether concepts should be seen as mental entities, representations, or linguistic entities (or else). Focusing on meanings of words has the advantage of avoiding such complications. Also, meanings can be rigorously defined in terms of intensions and extensions, and it is much easier to build a theory from such rigorous notions rather than on the nebulous notion of "concept".

So, Cappelen builds his account of Conceptual Engineering (or Meaning Engineering) from a purely linguistic standpoint. Then, when confronted with the question of how such linguistic modifications relate to modifications on the object-level, he argues that

Conceptual Engineering has a "**Worldly Characterization**" (Cappelen, 2018, p. 138). Using the example of the word "family", here is what he says:

The result is that we can say truly that, e.g., families have changed (and we can see that change as a result of conceptual engineering). What has changed is not the English words or some kind of abstract object called the concept of 'family'. Families are worldly phenomena and they have changed. (Cappelen, 2018, p. 140)

From this quote we understand that Cappelen thinks that modifications in the extensions and intensions of meanings have, as a result, a change in the object-level: changing words results in changing worldly phenomena as well.

2.2 Greenough's Objection

In response to Cappelen's idea about the Worldly Characterization of Conceptual Engineering, Patrick Greenough proposes an interesting tweak of perspective. In a forthcoming paper (Greenough, forthcoming) he suggests that we might want to look at Conceptual Engineering as a by-product of worldly modifications, thereby reversing Cappelen's order of priority. In order to argue for this position, Greenough summarizes Cappelen's view about the Worldly Characterization of Conceptual Engineering with the following principle:

"**BRIDGE**: For any meaningful term "*T*", this term undergoes revision of meaning if and only if *what T is (or what a T is)* undergoes revision." (Greenough, Conceptual Engineering via Reality Engineering, forthcoming, p. 3)

This principle formalizes the relationship between linguistic modifications and objectual modifications, saying that they both imply each other. From what we said above, we can see that Cappelen's view subscribes to the left-to-right direction (in the sense that he subscribes to only one of the two implication relations) of the principle: linguistic

modifications (in terms of intensions and extensions) imply objectual modifications, and not the other way around. It is relevant that the implication relation which is at work here is a rather robust one: under Cappelen's account linguistic modifications *cause* (or *ground*) objectual ones; it is not just a matter of one explaining the other.

Greenough recognizes relevant problems with the committal to the left-to-right direction of the principle, which, other than in Cappelen, is tacitly assumed in a number of views in the Conceptual Engineering debate. Indeed, it is understandable that Conceptual Engineers may be tempted by the left-to-right direction of BRIDGE: if Conceptual Engineering causes relevant objectual modifications, then Conceptual Engineering would acquire extreme theoretical and practical significance.

However. Greenough observes that the left-to-right direction of BRIDGE is "committed to a kind of linguistic construction of the world" (Greenough, Conceptual Engineering via Reality Engineering, forthcoming, p. 22), where a change in the referential apparatus automatically implies a change in the nature of things. As an instance: "we get to change the nature of women merely in virtue of revising the meaning of the word "woman"" (Greenough, Conceptual Engineering via Reality Engineering, forthcoming, p. 22).

As we observed above, the left-to-right direction of BRIDGE gives a strong metaphysical role to language, which is taken to *ground* certain entities. This consequence, however, is highly doubtable. Not only the relation of grounding between the word "woman" and the entity *woman* is debatable (more on this later) from a metaphysical standpoint, but it is clearly implausible to argue that a modification of the meaning of the word "woman" would be enough to modify the social status of women.

With this I do not want to say that linguistic modifications are generally irrelevant to objectual modifications, but their role should definitely be reshaped.

In section 3, I will analyse Greenough's positive proposal to reverse the order of priority of the BRIDGE principle. I am going to explore the implication of the committal to the right-to-left direction of the principle, thereby claiming that objectual modifications imply linguistic modifications, and not the other way around. The result of such committal is the formulation of an alternative movement to Conceptual Engineering, which we will call ***Reality Engineering*** (following Greenough).

Although I think that Greenough's argument is well constructed and that it represents a good reason to explore alternative options to Conceptual Engineering, it could be objected that such argument was built out a niche position (Cappelen's) which is but one of many in a rich debate. Also, one might be puzzled by Greenough's Uniformity Assumption, according to which all terms behave in the same way, in the sense that we can change "salad" in the same way we change "truth". To be fair to Greenough, he is the first one to highlight this problematic feature, but he nonetheless works with it. The general point is that Conceptual Engineers usually do not distinguish between different kinds of terms, which is troublesome.

Therefore, I will now provide a second argument against the sufficiency of Conceptual Engineering with respect to its own goals of changing the world. In particular, I will focus on the relationship between Conceptual Engineering and the social World, since the social realm is going to be the theme of the whole dissertation.

2.3 Conceptual Engineering and the Social World

The Conceptual Engineering literature sparked within theoretical philosophy, and interesting projects of conceptual modifications have been developed with respect to

such theoretical concepts, like Scharp's discussion about *truth* (Scharp, 2013) and Clark and Chalmers' discussion, although not explicitly in engineering terms, about *belief* (Clark and Chalmers, 1998).

Many philosophers, however, expanded the scope of the discipline, focusing also on social and moral concepts. About social Conceptual Engineering, the most prominent example is the one of Sally Haslanger's discussion about *gender* (Haslanger, 2005, 2006 and 2000), followed by Esa Diaz-Leon (Diaz-Leon, forthcoming), Amie Thomasson (Thomasson 2020), Teresa Marques (Marques, 2020) and Manuel Gustavo Isaac (Gustavo Isaac, 2021).

The shared position of these philosophers is that some features of social Reality are defective, and we should try to change them. These philosophers individuate Conceptual Engineering as a tool to achieve such modifications and, somewhat similarly to Cappelen, there is little to no discussion of modifications out of the linguistic realm.

An interesting consideration about the reasons and goals of social Conceptual Engineering is provided by Manuel Gustavo Isaac (2021), in response to Mona Simion's controversial position (Simion, 2018). In her paper from 2018, Simion argues that Conceptual Engineering should always be motivated by epistemic reasons and goals: we should modify concepts when they are epistemically defective in order to make them epistemically more adequate.

On the contrary, Gustavo Isaac 2021, p. 5) argues that epistemic reasons are not the right ones to follow in the case of social Conceptual Engineering. In his framework, this consideration follows from the constructivist claim that social concepts are ideological (as opposed to natural), and in terms of Podosky (2018, p. 6) "ideologies make

themselves true". Therefore, in order to deconstruct ideologies, we should not follow epistemic reasons (since social concepts conceived this way justify themselves on an epistemic standpoint). Here is how he puts it:

We want some representationally inaccurate (and therefore truth-unconducive, so to speak) concept to replace/revise some representationally accurate (and therefore truth-conducive) concept. And provided that the resulting social change is taken to be good (e.g. as promoting social justice, etc.), such epistemic loss is expected to be beneficial. (Gustavo Isaac, 2021, p. 5)

Leaving aside the discussion about ideologies, I think that Gustavo Isaac has a great point here, which is worth highlighting:

- **Social Conceptual Engineering** should be driven by the will of generating amelioration in social Reality in terms of social justice, wellbeing, and other similar reasons. The point is to change the world, not just the concepts.

The lesson to extrapolate here is the following: social Conceptual Engineering is a unique sub-discipline of Conceptual Engineering, and it has its own reasons and goals. Therefore, it is worth to analyse this sub-discipline in isolation, without conflating the considerations about it within the non-specific discussion of Conceptual Engineering.

However, also under Gustavo Isaac's position there seems to be an implicit assumption that conceptual modifications will be able, by themselves, to generate relevant modifications in the social realm. My goal is to show that this is not the case. I will argue for this by showing that social Conceptual Engineering is committed to social Constructivist metaphysical views. Then, I will analyse the relation between concepts and entities, under Constructivist frameworks. Then, I will formulate an Insufficiency

Challenge, which is, in my opinion, a hard issue to overcome for Conceptual Engineering.

2.4 Social constructivism as a metaphysics for Social Conceptual Engineering

The literature about social metaphysics is extremely wide, and there are different ways to taxonomize it. I think that the best way to do it is to distinguish between three patterns, under which we can group most of the views. Surely this division will not result exhaustive of all the views about this topic, but it allows us to grasp some key distinctions.

The three patterns of views in social metaphysics are (in my own terms) *social naturalism*, *social constructivism* and *social eliminativism*. In simple terms, social naturalism represents the group of views that say that social entities exist, and their existence is mind-independent. Social constructivism groups those views that claim that social entities exist, and their existence is mind-dependent (in the sense that they depend on our beliefs, attitudes, practices, behaviours, linguistic uses, and so on). Social eliminativism collects positions that claim that social entities do not exist at all.

Nowadays, social naturalism has almost entirely been dismissed. Such framework runs in deep trouble due to some epistemic challenges (if social entities exist mind-independently, how do we get to know them? How can we claim that the social categories that we actually live by coincide with the *real* social categories?) and some moral ones (the risk is to justify unjust social mechanisms with the claim that they are naturally grounded). But even regardless of these theoretical problems, we can safely dismiss social naturalism as a metaphysical view for Conceptual Engineering. If we grant that there are some *real* social categories out there, and that the defectiveness in our social world is generated by bad concepts that fail to represent such *real* categories,

then Conceptual Engineering's aim would be solely an epistemic one. The goal would be to modify concepts in order to make them *represent reality better*. As we have seen in the discussion between Gustavo Isaac and Simion, this approach is untenable. In social Conceptual Engineering, we do not care about adequate representation, we care about making the (linguistic) world a better place.

social constructivism is, on the other hand, a valid metaphysical framework for Conceptual Engineering. By arguing that social entities exist, it provides a good theoretical background: the world that Conceptual Engineering is trying to change is ontologically strong. By arguing that the existence of social entities is mind-dependent (in the sense specified above), it provides a good practical standpoint: what we think and do really matters for shaping the social world. I will dig deep into different views in the social Constructivist literature later on in the dissertation, and I will analyse the relationship between social entities and social concepts under the Constructivist framework in the next section. But first we need to argue against social eliminativism as a metaphysical framework for social Conceptual Engineering.

social eliminativism says that social entities (at least some of them) do not exist at all. Usually, such metaphysical position comes from a normative proposal to eliminate certain social entities. For instance, some philosophers that aim to eliminate genders from society, like Wittig (Wittig, 1992), think that some elimination is grounded on the fact that genders do not exist. However, this jump from the practical will to eliminate genders (or races, like in the case of Appiah) to the metaphysical claim that they do not exist is far from being necessary. Social constructivism is a valid alternative, under which we can change (or get rid) of things maintaining their ontological status. Such jump is, I believe, caused by a misunderstanding. Claiming that social entities exist does

not mean that they exist mind-independently. It does not mean that they exist naturally. Claiming that they exist just follows from the observation that we live by some social categories, and the fact that we recognize them, respect them, use them, criticize them is a proof of it. Sally Haslanger is a champion of this approach. According to her, gender exists as the result of oppressive social mechanisms, and we should get rid of it (Haslanger 2005, 2006). From a strictly metaphysical standpoint, a good response to social eliminativism can be found in Amie Thomasson's *easy* approach (Thomasson, 2015). According to Thomasson, those who deny that social groups exist cannot refer to them and thus cannot speak of them in ordinary discourse. (Thomasson 2019, 4833). Social entities are useful for us to navigate the world, whether we like them or not, and this is enough to claim that they exist under the *easy* approach. I will defend this metaphysics of social entities more accurately later on.

For now, however, it is enough to see that social eliminativism is not a good view in social metaphysics. And even if it was, it is not well-suited for social Conceptual Engineering. Social Conceptual Engineers does not only aim at the destruction of our social Concepts (like eliminativists want) but also the creation of new, better concepts. This positive part of the proposal is incompatible with the Eliminativist framework.

So, we can conclude that social Conceptual Engineering is (or should be) committed to social constructivism as a metaphysical view.

2.5 Social Concepts and Social Entities

To recapitulate, my goal here is to argue that Conceptual Engineering cannot, by itself, fulfil its goal of changing the social world for the better. The components of my argument are the commitment of social Conceptual Engineering to social constructivism and the analysis of the role of concepts in Constructivist frameworks. With these two

elements in place, we will be able to formulate the Insufficiency Challenge. So, this section is dedicated to the analysis of the relationship between social concepts and social entities, under constructivism.

Given, from constructivism, that social entities are mind-dependent and constructed by us in some way, we could ask what exactly it is that constructs them. I will dedicate long discussions to some of the many views that attempted to analyse this construction process, but for now I will remain on more general grounds.

There are two options about the role of concepts in such construction: either concepts do not contribute to the construction of social entities (a view that we can call Representationalism) or they do (on a view that we can call normativism). We can immediately see that this distinction is unbalanced, since normativism includes a broad spectrum of degrees in which concepts contribute to the construction of social entities. Representationalism is, therefore, a rather extreme view in this spectrum since it is the view that concepts do not contribute at all.

normativism has a plethora of *mild* forms, and one extreme form, which is that social concepts are the sole contributors to the construction of social entities. The most sensible versions of *mild* normativism claim that social entities are constructed via uses, practices, behaviours, act, beliefs, attitudes, *and* linguistic utterances. So, one might have a *conjunctivist* reading, that puts all these components of construction on a par (from a priority standpoint), or a *scaled* reading that says, for instance, that linguistic utterances and concepts ground our practices (or vice versa).

The key point here is that whatever account of the role of concepts in the construction of social entities you take, except the implausible extreme normativist view, you end up with an Insufficiency Challenge against social Conceptual Engineering.

2.6 The Insufficiency Challenge

In order to formulate the Insufficiency Challenge, I will break down the discussion in three parts. First, I am going to show that Extreme Representationalism leads to obvious problems. Then, I will discuss problems with Mild normativist views. Lastly, I will show that Extreme normativism, although it is resistant to Insufficiency issues, untenable on different grounds.

The argument for the Insufficiency of social Conceptual Engineering with respect to its goal to effectively modify the social world is quite obvious if we work under an Extreme Representationalist account of the relationship between social concepts and social entities. If concepts are just representations of social entities and they do not contribute at all in their construction, then the modification of such concepts will not bring any effect in terms of changing such social entities. Even further, under Representationalism, social Conceptual Engineering makes sense only in Simion's terms of making concepts better epistemically. You have fixed social entities and you try to make your concepts represent them more adequately. This, as we saw with Gustavo Isaac's contribution, is not what we want social Conceptual Engineering to do. Changing social concepts for epistemic reasons does not help us changing the defectiveness of the social entities that are at work. If anything, it justifies such defectiveness.

The discussion about *mild* normativism is more complicated. *Mild* normativist positions hold that social entities are constructed by uses, practices, behaviours, act, beliefs, attitudes, *and* linguistic utterances.

On a *conjunctivist* reading of *mild* normativism, it is easy to see that social Conceptual Engineering would be insufficient in order to modify social entities in their nature, since, together with concepts and linguistic devices, a change in practices is also required. In this case, social Conceptual Engineering is necessary but not sufficient for the modification of the social world.

On a *scaled* reading of *mild* normativism where practices ground concepts the Insufficiency problem is also evident. Changing practices would imply a conceptual change, and not the other way around. Therefore, social Conceptual Engineering would be neither necessary nor sufficient for the modification of the social world.

On a *scaled* reading of *mild* normativism where concepts ground practices, it seems, at first sight, that Insufficiency does not appear. However, we have to be a little careful. If concepts are taken to be the *only* things that ground our social practices, then Insufficiency would not arise, but that would not be *mild* normativism anymore, since concepts would be, although indirectly, the only contributors to the construction of social entities. And, as we will see in a moment, extreme normativism is not an appealing view to have. So, we would have to say that, although concepts ground our social practices, some other things do too. I admit that this view does not sound very plausible, since it does not even specify what else contributes to the grounding of our social practices, but this is far from the point. The point is, once again, that changing concepts would not be enough, by itself, to change our social practices. Hence, it would not be enough to change the social world.

Lastly, we have to address the extreme version of normativism. Surely, under such an extreme view where social concepts are the sole contributors to the construction of social entities, or the sole contributors to the grounding of our social practices,

Insufficiency challenges would not arise. Changing concepts, by itself, would lead to a modification of the social world. However, this view is extremely committing, and to be fair, quite implausible. An accurate breakdown of this view would require an equally accurate breakdown of the nature of concepts, which I will not attempt here. So, I will not be able to completely dismiss this view. But what I can do is show that the view is unappealing. What this view is saying is that all our social behaviours, uses, attitudes, beliefs are caused by or grounded on linguistic and representational devices. This seems to be a rather reductionist and simplistic way to characterize our social practices. How would linguistic devices ground our behaviours? If social concepts are prior to social practices, where do they come from? These would be questions to answer for someone that wants to defend the Sufficiency of social Conceptual Engineering with respect to modifications of the social world. I do not think that such answers can be found easily, and I am afraid that the price in terms of commitments that one would need to defend such positions are much greater than the benefit.

In conclusion, I would like to highlight that this extreme version of normativism seems to be the tacit assumption of the vast majority of social Conceptual Engineering projects. Social Conceptual Engineering believes (either explicitly, like Cappelen, or implicitly, like Gustavo Isaac) that the modification of linguistic/referential/representational devices can bring significant modifications to the entities of the social world. It would be interesting if Conceptual Engineers dedicated more attention to this issue, since I think that it would cause a shift in the focus of the discipline. Indeed, such shift of focus is my own goal. I believe that Conceptual Engineering is insufficient with respect to its own goal to modify the social world, and I

want to analyse other, more worldly oriented, ways to impact such objectual modifications.

2.7 The Origin of Implementation Problems

In conclusion of this chapter, I just want to point out an interesting idea. Without a doubt, the most discussed problem in the Conceptual Engineering literature is the so-called Implementation Problem. In easy terms, the question is: Even if we successfully modify a concept, how do we get people to abandon the old bad concept in favour of our new better one? How do we apply this conceptual modification to the world?

I think that the origin of this problem lies exactly in the discussion that I provided in this chapter. Conceptual Engineers have been way too optimistic about the role of their discipline: they took for granted that we can change concepts without modifying the practices that ground them. Surely, this can be done within small philosophical communities, where everyone shares the same attitudes and beliefs about a certain topic. Within our philosophical practices is, for instance, easy to see the treatment of gender and race in society is deeply unjust. We can therefore develop better concepts of gender and race for us to work with. But that is not, by itself, going to change people's beliefs, or policies (and politics). Conceptual work is not going to change anyone's status in society.

This, I think, is due to the fact that conceptual and linguistic devices are shaped by a deep network of social interactions that cannot be changed with a snap of our fingers. Implementation of conceptual modifications is bound to fail if we keep ignoring that we have to change people's mind before we get to change the way they speak.

Again, I take this discussion to be a compelling reason to look at alternative, more worldly oriented, versions of Conceptual Engineering, such that conceptual work will become parallel, if not subordinate, to concrete actions about our practices.

Chapter 3 – Reality Engineering as an alternative project

3.1 Reasons to look for an alternative account

From the last chapter, we can extrapolate compelling reasons to explore alternative accounts to Conceptual Engineering, without ignoring some of its good features. The point, at least for what concerns social Conceptual Engineering, is that the ultimate goal of engineering projects is to not only modify linguistic and/or representational devices but to also modify the world we live in, trying to make it a better place. However, we showed that changing linguistic and/or representational devices does not, by itself, entail a modification of the social entities and categories that we live by. This is due to a peculiar metaphysics to which Conceptual Engineering seems to be committed, i.e. social constructivism. Under such metaphysics, we have seen that whatever view on concepts we might want to commit to, either Representationalism or the various forms of normativism, we end up with either an untenable view or an Insufficiency Challenge.

This specific argument, together with Greenough's criticism of the Conceptual Engineering discipline in general, shows us the need to look at some alternative account, more worldly oriented, in order to accomplish the goal of making our social world better.

The point is, therefore, to find an alternative which aims to modify *things* in the world, rather than mere concepts. If we take a look once again at Greenough's BRIDGE principle, this will appear clearer:

"BRIDGE: For any meaningful term "*T*", this term undergoes revision of meaning if and only if *what T is (or what a T is)* undergoes revision." (Greenough, Conceptual Engineering via Reality Engineering, forthcoming, p. 3)

As we pointed out, Conceptual Engineering views usually subscribe only to the left-to-right direction of the principle. This means that they believe that revision of meanings (or of concepts) causes the revision of the nature of the thing that such meanings (or concepts) refer to. After having criticised this view, my goal, similarly to Greenough, is to explore the reversal of this order of priority.

So, by refuting the view that conceptual modifications imply worldly modifications, which Greenough calls "Representation-First View" (Greenough, forthcoming, p. 4), we end up with either a position that holds that conceptual modifications and worldly modifications happen at the same time or a position that claims that worldly modifications imply conceptual ones. In Greenough's words, the first one is called "Equi-Basic View" (ibid, p.5) and the second "Reality-First View" (ibid, p.4). At this early stage it is not relevant to argue for one of the two instead of the other, what is worth noticing is that both require an account of what worldly modifications are and how to make them happen. To give it a name, they both need an account of *Reality Engineering*.

3.2 Two Big Questions

The reader might fund the label "Reality Engineering" puzzling since making sense of the expression "changing reality" in philosophical terms is far from being an obvious task. Out of the many questions that can be asked in this preliminary stage, two appear to be particularly relevant: What does it mean to change Reality? Should we even do it?

The first is the question that will occupy the entirety of this dissertation, while, for now, I have to set the second one aside. The latter question is, in fact, a moral, political, normative question. There are two reasons why I will not deal with the issue here: the first one is that it would open an overwhelmingly big discussion, which is also relatable

to Conceptual Engineering in its original form, and that would add complications that I would not be able to analyse properly within this dissertation; the second reason is that such moral conundrums are especially applicable to individual engineering projects, rather than on the discipline as a whole, and for the purposes of this work I will remain on general grounds, without attempting to formulate any individual instance of Reality Engineering.

Therefore, I will dedicate my attention to the question of what it means to change reality. In order to provide a proper analysis, I will break down the question in two parts: the Scope issue and the Manual issue.

The Scope issue concerns the subject matter of Reality Engineering: what parts of reality we are trying to change. The only other account of Reality Engineering, which is the one developed by Greenough, considers properties to be the scope of the project. According to Greenough, Reality Engineering should focus on the modification of properties that things possess, like the property of "being woman", "being salad" and so on. I think that Greenough's proposal is metaphysically vague on the explanation of the nature of properties, and not really clear in that regard. In the next section, I will make the case that Reality Engineering should take kinds as its scope. As we will see, there might be ways to translate property-talk to kind-talk and viceversa, but I think that kind-talk is more understandable and metaphysically promising. This applies especially to social Reality Engineering, which is the topic of the dissertation, where it is more helpful to talk about social groups and categories rather than social properties.

The Manual issue concerns the way to proceed in Reality Engineering projects. The Manual issue is, of course, heavily dependent on the Scope one. How to change worldly things depends on what we are trying to change. This is the reason why I will dedicate

much discussion to different views about the nature of social categories. If we set an accurate metaphysics of social categories, then it will be much easier to understand what we can do to change them. On general pre-theoretic grounds we can say that, oppositely to Conceptual Engineering, the methods of Reality Engineering are going to be practice oriented, with strong social and political connotations.

The two issues combined show that Reality Engineering is nothing else than a theory of social Action (Manual), enriched by guidelines about how to direct such Action (Scope).

3.3 Reality Engineering in History and in Everyday Life

Before we get in the discussion about the philosophical technicalities of Reality Engineering, I want to point out that such discipline is nothing new. Projects of modifications of social entities happen on a daily basis in our society, and so has been the case in the past.

We can see countless instances of projects of modifications of social entities and mechanisms. Just to quote a few, we can read the anti-racist Black Lives Matter movement as a project that aims to modify the social category of people of colour with the goal of improving their social status (in broad terms). But we could also look at Feminist movements, LGBTQ+ movements, Pro-Rights movements as contemporary examples. Also, we can read historical facts such as the Russian Revolution, or the Italian Resistance movement against fascism under this light. All of these examples have in common the will of modifying some defective features of the social world, which is the defining goal of social Reality Engineering.

The point is that Reality Engineering has already happened in many circumstances, and it still happens today on a daily basis. What my transposition of these matters in

philosophical terms is trying to do is to provide a robust philosophical background for projects of social Action, with the hope that philosophers and members of various communities will work together to change the social world for the better.

One might be puzzled by the fact that social Reality Engineering, which claims to be a philosophical movement, is, in the end, not philosophical at all. I hope to show that this is a positive feature of this account. Philosophers cannot, by themselves, change the world, but they can contribute.

Within the many points that I make in this dissertation, this metaphilosophical claim is, in my opinion, the most important: philosophers should step down from their armchair and actively engage with the world, using their philosophical training to collaborate with social activists, politicians, and various members of various communities.

Chapter 4 – The Question of Scope: Kinds vs Properties

4.1. Greenough's Proposal

The question of Scope is extremely relevant for the Reality Engineering movement. This follows from the idea that an accurate understanding of what exactly we are trying to change is fundamental in order to change it effectively. What we want, therefore, is to identify the *portion of Reality* that we are trying to change.

A preliminary point to make is the justification of why I limit my search for candidates to be the scope of the discipline to philosophically relevant entities only. Again, my interest is to analyse deeply what we are trying to change. I want to do it by analysing the metaphysics of the entities that we will pick as scope. Much metaphysical work has already been done to analyse features of different entities, like objects, properties, categories, and many more. I will rely on work that has already been done, and I will pick the kind of entities that I believe has been analysed in the clearest way by the literature.

Let us then start by analysing what Patrick Greenough has said about this topic.

In his paper, Greenough talks about both kinds and properties. He is not fully precise in the definition of neither kinds nor properties, but both seem to make his argument work. When first presenting the issue, working with the example of *being a computer*, here is how he puts it:

"In effect, *being a computer* is a *plastic property*. The plasticity of the meaning of "computer" thus goes hand-in-hand with the plasticity of the property of *being a computer*. Furthermore, if the things which have this property are sufficiently unified to form a kind (in virtue of having this property), then we can also recognise *plastic kinds*." (Greenough, forthcoming, p.21)

This quote shows that Greenough's primary interest is on properties: properties are plastic (i.e. they change over time), and the change in properties is what we deal with in Reality Engineering. Then, he says, we can extend the discourse to kinds. This can be done by individuating a group of things that share the relevant property: since the property is plastic, the kind is plastic too.

I think that Greenough here is deliberately non-technical, the point is about plasticity and change over time, not on properties and kinds. But for the reasons above, I cannot have the luxury of not picking. Therefore, I pick kinds. Below I show that there are multiple reasons to pick kinds over properties.

4.2 Kinds vs Properties

I will here lay out four points to deal with the issue of kinds over properties:

- *Property-Talk and Kind-Talk are not equivalent.*

  First, I admit that property-talk and kind-talk can easily be translated into each other: you can build a kind out of any property (as Greenough does), and you can dismiss kinds just by looking at things that share properties. Fair enough, but Greenough's move shows that the kinds that we end up with are not exactly those that we would want. Using Greenough's trick, it seems like we end up with kinds that are identified by a single property. For instance, the kind *computer* is identified by the single property *being a computer*. From the literature, however, we know that kinds that are identified by single properties are rather uninteresting. As an example, we can take Boyd's Homeostatic property Cluster View about natural kinds (Boyd, 1999), where kinds are identified by multiple properties entrenched indifferent ways. Also, most of the literature that concerns social entities focuses on kinds, not properties. We will

see this thoroughly in Chapter 5. One potential response could be that the property *being a computer* is some kind of second-order property, built out of more basic properties like *having a software*, *having a hardware*, *having a screen*, and so on. I am not fond of this hierarchic way to look at properties, and I find it, at most, unnecessarily complicated. The complications brought by this hierarchic move are avoidable since we can gain the same positive aspects with the much easier metaphysics of kinds. So, to recapitulate, the first reason to prefer kinds to properties is that the metaphysics of kinds is easier to understand and more analysed by the literature that concerns our topics of interest.

- *Plasticity works better with Kinds.*

  The second reason to prefer kinds to properties is the discussion of plasticity (i.e. change over time) and Engineerability. Greenough argues that properties are plastic entities, in the sense that they can change over time while maintaining their identity. As we have seen, his examples concern properties like *being a computer*, and it is easy to see that such a property can indeed change over time. As Greenough rightly observes, something that counted as a computer twenty years ago would not count as a computer today, and what will count as a computer in twenty years does not count as a computer today. It is intuitive, therefore, to say that the property *being a computer* changes over time. I am afraid, however, that such intuitiveness is lost if we look at different properties, such as, for instance, *having the atomic number of 7*, or *being brown*. *Having the atomic number of 7* is a property that applies to chemical elements (Nitrogen). Surely enough, the number of Nitrogen molecules in the world changes over

time. So one could say that the property of *having the atomic number of 7* changes over time, since the amount of things possessing such property has changed. This, however, does not seem to be the way in which Greenough understands plasticity and change. What counted as a molecule *having the atomic number of 7* centuries ago would still count as *having the atomic number of* 7 today, and always will. In technical terms, the property of *having the atomic number of 7* can only change on an *extensional level*, i.e. can change in terms of which things possess the property. However, such property cannot change on an *intensional level*, which means that what it takes for something to possess the property does not change. We are going to discuss the meaning and role of Intensions and Extensions later on, but for now it is enough to observe that Greenough is interested in what we can call Intensional plasticity. This means that Greenough looks at properties that change in terms of the conditions that things have to meet in order to count as possessors of them. Many properties do not seem to exhibit such Intensional plasticity. On the contrary, kinds are more compatible with the idea of Intensional plasticity. We can see this even under a non-so-technical conception of kinds such as the one that we have now. As we have seen above, kinds should not be considered as identified by a single property nor as second-order properties. It is theoretically and practically advantageous to think about kinds as collections of entities that share clusters of properties. Such properties can be linked in different ways, this should not concern us at the moment. The point is that, under this light, Intensions of kinds are the conditions that something must meet in order to count as a member of the kind, i.e. Intensions are definitions of kinds in terms of the properties that things

must possess in order to gain membership to the kind. Therefore, Intensional plasticity for kinds concerns the possibility of the variation of which properties are relevant to gain membership to a kind. This does not imply that the nature of single properties has to change, but just that things can acquire or lose possession of certain properties, and this may lead to the gain or loss of membership to a certain kind. Therefore, kinds seem to have the kind of plasticity that Greenough is interested in, but properties do not. I share Greenough's idea that Intensional plasticity is the form of plasticity that we should care about. Hence, these considerations about the better behaviour of kinds with respect to Intensional plasticity count for me as a reason to adopt kinds as the Scope of Reality Engineering.

- *No Uniformity Assumptions*

The third reason to prefer kinds to properties is the rejection of the controversial Uniformity Assumption used by Greenough. In his paper, Greenough works under the so-called Uniformity Assumption:

> "exhibit a clear case where some (general or mass) term (e.g. "salad") has changed in meaning; make the case that this case is representative of a range of terms in natural language, including philosophical terms; and so, conclude that all terms of philosophical interest can (and do) change in meaning in just the same general way as the exhibited term." (Greenough, forthcoming, p.26)

In this quote, Greenough states the Uniformity Assumption for what concerns meanings. He assumes that all terms behave in the same way. Under such

assumption, if some terms can change in meaning, then all terms can change in the same way.

He makes a similar assumption about properties: if some properties can change over time while maintaining their identity, then all properties can change in the same way. This clarifies the reason why Greenough chooses such general instances of properties that change over time (like *being a computer*). It suffices for him to find a few instances to conclude that all properties can change, given the Uniformity Assumption. Of course, we should not take Greenough's proposal as an argument for the validity of the Uniformity Assumption, we should just take it as a working hypothesis. Such Assumption makes sense in Greenough's framework, because he is just trying to provide a general guideline for the project of Reality Engineering, and he has no intention of developing it in detail in the paper. Again, unfortunately, I have to renounce to the luxury of working under Greenough's framework, since I am indeed trying to give shape to Reality Engineering. I cannot accept the Uniformity Assumption. Also, I see no way to argue for the truth of the Assumption. Although not explicitly, I already argued against the Uniformity Assumption in the previous point: I showed that some properties change intensionally and others do not. Therefore, we can discharge the claim that all properties change in the same way. Now, this point is relevant because by denying that all properties change in the same way we commit to providing a differentiation of properties. What we would need is a metaphysics that distinguishes between different kinds of properties and explains which of them change and in which way they change.

There are instances in the literature that concern discussions around natural properties (like in (Lewis, 1983) and (Hirsch, 1993)). Much focus has been put on how to recognize properties that are *more natural* than others, but little has been said about what non-natural properties are and how we should understand them. My interest is in the study of social entities, and almost no one amongst social Metaphysicians has engaged with the analysis of social properties. Much more discussion has been dedicated to entities like social groups, Categories, kinds, and so on. Therefore, I consider more practically useful to focus on social kinds, so that I will be able to engage with the literature in a more direct way.

- *Different kinds of kinds*

In conclusion, I will add a couple of remarks about the idea of differentiating between kinds of kinds. Just to give the reader the sense of how tricky it can be to divide kinds, I present a partition proposed by Kerr, in a paper from 2014 (which inherits traits from (Kusch, 1999)). Kerr distinguishes three kinds of kinds: natural kinds, social kinds and artificial kinds.

Kerr takes natural kinds to be "wholly identified by their empirical characteristics" (Kerr, 2014, p. 211) (examples *Calcium* and *Cheetah*). This means that natural kinds are entities identified only via worldly features, no human action enters the process of their definition.

Secondly, he thinks that social kinds can "be reduced to an entirely social system of fiat" (Kerr, 2014, p. 212) (examples *Marriage* and *Law*). By this, Kerr means that social kinds are determined by social practices and relations.

Last, artificial kinds are a mixture of the two above, they are identified via empirical characteristics and social practices (examples *Hammer*, *Microscope*, tools in general).

Kerr's proposal is an instance of a non-typical partition, since not many philosophers consider social kinds and artificial kinds to be distinct, and neither do I. This, however, shows that there is no fine-tuned robust distinction between different kinds of kinds, and we really should take the definition of a social kind in the most general way possible. My personal way to identify social kinds is to merge Kerr's definition with the one of artificial kind, so that all those categories that would not exist if it were not for the presence of our social practices count as social kinds.

I will dedicate the following long chapter to a more detailed discussion of the metaphysics of social kinds.

Chapter 5 – A Taxonomy of the metaphysics of Social Kinds

Chapter 5.1 – The Constitution Account

The first position in the metaphysics of social kinds that I am going to present is the Constitution Account. The origin of this view can be found in the very influential book by John Searle "The Construction of Social Reality" (Searle, 1995). The account was later developed by Amie Thomasson, in two papers from 2003, "Foundations for a social ontology" and "realism and Human Kinds". Since Thomasson provides interesting insight and additions on Searle's original theory, I will mainly refer to her two papers for the present discussion.

I am going to proceed as follows: I will start off by showing Searle's original theory, highlighting his main claims. Then, I am going to discuss Amie Thomasson's contributions, by analysing the issues of the relationship between social kinds and physical objects, the contrast between this theory and realism (or social naturalism), and the epistemic challenges related to the role of agents, error, and transparency. I will conclude the section by highlighting some potential problems with the view.

5.1.1 Searle's social metaphysics

So, in his book from 1995, Searle argues that the construction of social reality has three pillars: "The assignment of function, collective intentionality, and constitutive rules" (Searle, 1995, p. 13).

Searle's idea is that we create social facts by imposing a function on physical objects (that can be pre-existent or created for such purpose) such that those objects could not fulfil the function solely in virtue of their physical composition. Furthermore, the function cannot be assigned arbitrarily by a few agents but must be collectively accepted by all (or most) the members of a relevant community.

In the case of institutional kinds (which is just a different label for social kinds), the relevant function must be expressed in terms of constitutive rules. These rules stipulate that "a certain x "counts as y in the relevant context C" (Searle, 1995, p. 44). In this definition, x stands for some physical object, upon which we impose the social function y, which is only relevant within a certain context (or community) C. In order to generate a proper social kind, these constitutive must be collectively accepted by the members of community C.

According to Searle, even though social kinds are ontologically subjective (given that we intend acceptance as a result of a mental state), they are epistemically objective: the fact that their existence is grounded on collective intentionality implies that their truth or falsity is independent of a single epistemic agent's beliefs or concepts.

5.1.2 Thomasson on the relationship between social kinds and physical objects

In Searle's framework, social Entities require physical objects as the base upon which we impose a peculiar function. For him, the physical world has the ontological and logical priority over the social world: without a physical substratum, no social entity can exist.

Thomasson, however, does not agree with this idea. In "Foundations for a social ontology", she turns Searle's main point against him. Searle took the priority of the physical as a good feature of his own account, since he was afraid that a social reality detached from the physical world would suffer from a low ontological status: it would not be considered *real* enough.

The problem, according to Thomasson is that Searle's account seems to cover only a few social kinds. A relevantly big class of social facts and entities remains left out. Not all social entities have a physical substratum.

As instances of these left out entities she takes "the U.S. Constitution, General Motors, or the Calvinist doctrine" (Thomasson, 2003 (a), p. 273). About these entities she says that "If we are to take the[m] […] at face value, rather than attempting to find some way to paraphrase all talk about them into talk about concrete individuals, we will have to admit that new social *objects,* not just facts, can be created" (Thomasson, Foundations for a social ontology, 2003 (a), p. 273).

This quote is particularly interesting because I think that this is the core of her motivation to move away from Searle's ontology and such motivation is not entirely uncontroversial. Redescriptions are a typical strategy in metaphysics, from the antirealist side. Since according to Quinean metaontology the ontological value of an entity is given by the possibility of that entity to be the value of a variable, and we should be committed to all and only those entities whose reference is necessary for our explanation, the antirealist usually tries to show that we can provide the same explanation without referring to a certain kind of entities, which existence they want to deny.

For instance, an antirealist about numbers will try to paraphrase mathematics in order for it not to contain any instance of number-talk, and still achieving the same results.

Amie Thomasson, however, does not subscribe to this strategy, as can be observed in her later work "ontology made easy", from 2015. Therefore, she is inclined to discard the option of redescribing things and rather prefers to take them at their "face value" (Thomasson, 2003 (a), p.273).

Of course, since her main work on ontology is from 2015, and this paper from 2003, in the latter she could not put forward the good arguments that she makes in the former. So, here I will try to back up the Thomasson of 2003 with her later contributions.

In her book from 2015, Thomasson argues that ontological questions can be answered in *easy* straightforward ways, with the help of nothing else than "conceptual and (/or) empirical investigation" (Thomasson, 2015, p.11). The point is that we all possess the concept of social kind, and of individual social kinds. We also see that they are empirically useful for our social lives. This, according to Thomasson, is enough to claim that social kinds exist. Similar arguments are present in her work from 2019 as well, but we are going to discuss them together with the Normative Approach later on.

Therefore, the worry of having social kinds that are *not real enough* is, for Thomasson, completely misguided. Hence, we should not be so committed to the link between social kinds and physical objects. In this way, we can account for many more social entities than Searle could, including those that do not have a physical substratum. As long as the imposition of function, collective intentionality and constitutive rules are in place, we can account for the existence of social kinds. Of course, constitutive rules will take a slightly different form from Searle's. I am going to show this in a bit while presenting Thomasson's positive proposal. Now, however, it is time to look at another of her criticisms.

5.1.3 Thomasson on constructivism vs realism

The second criticism by Thomasson concerns the relationship between Searle's Constructivist account and realism about natural kinds. What Thomasson does here is to show that social constructivism is incompatible with the idea that social kinds are equal to natural kinds, in both metaphysical and epistemic regards. In terms of our initial discussion about constructivism, here Thomasson offers us a back-up argument against social naturalism (the view that social kinds exist mind-independently). This is relevant because Searle, as we have seen above, was hesitant about the relationship between

social kinds and the world since he feared that they would not be considered to be *real* enough.

Thomasson (2003 (b)) challenges the idea that realist accounts (in ontology, epistemology, and semantics) that were developed to accommodate natural kinds can be extended to cover social kinds as well.

The ontology of social kinds violates the ontological constraint of Mind-Independence that is a core component of realism for natural kinds. Interestingly, this ontological difference implies sensible differences for what concerns the epistemology of social kinds and how we refer to them.

In order to illustrate this, Thomasson starts off by delineating the three columns of realist accounts for natural kinds.

First, the ontological component. According to Thomasson, realism for natural kinds must be committed to an ontological principle, which she calls the Independence Principle:

- Independence Principle: "Things of kind K exist independently of the mental, […], it is possible that there are things that are of kind K and that there are no mental states whatsoever" (p. 582)

In other words, the realist position claims that certain things exist independently from our beliefs, concepts, attitudes, etc. One concern with the principle put this way is that it fails to imply that kinds exist independently from us, since it just claims that individual things do. Furthermore, it seems to me that the principle fails to imply that the things which exist independently from us are *naturally* things of kind K. It seems prima facie plausible that individual things exist independently from us and we distinguish them

using kinds which do not exist independently from us. Perhaps the realist account is not well-represented by this principle.

Also, according to the realist, the kind K has natural boundaries, which are not dependent on human beliefs or concepts.

From this ontological principle, epistemological considerations follow. These epistemological positions can be put under two principles: the Ignorance Principle and the Error Principle.

- Ignorance Principle: If the Independence Principle is true, then "for all conditions determining the nature of the kind K, it is possible that these remain unknown to everyone". (p. 583)
- Error Principle: If the Independence Principle is true, then "any beliefs regarding the nature of Ks could turn out to be massively wrong". (ibid.)

To explain, the Ignorance Principle says that we might not know which factors determine what things are of a certain kind K. From here follows the Error Principle, if we do not know which factors determine the boundaries of a kind, we could be very wrong about them.

Following these principles, Thomasson says that the realist account is inclined to accept Causal Theories of Reference for what concerns the semantic portion of the view. As she puts it, "causal theories of reference are based on the idea that there is a kind with pre-existing boundaries that can determine the extension of the term regardless of speaker's beliefs and concepts regarding the kind" (ibid).

That is to say that, for the realist, it is the world that shapes the boundaries of our language and not vice versa.

Now, social kinds fail to meet the Independence Principle: if K stands for a social kinds, it is not true that things of kind K exist independently of all mental states.

But it is in principle possible that some kind K fails to meet the Independence Principle but still meets the Ignorance Principle, the Error Principle and its reference does not violate causal theories of reference.

The goal of Thomasson's argument is to show that, in the case of social kinds, a failure to meet the Independence Principle implies a failure to meet the epistemological and semantic realist views as well.

As we know, for what concerns social metaphysics, Thomasson follows Searle. Recall that Searle says that institutional facts are those social facts such that we "collectively impose a function on a phenomenon whose physical composition is insufficient to guarantee the performance of the function, and therefore the function can only be performed as a matter of *collective intentionality* or *recognition*" (Searle, 1995, p. 124) (my emphasis).

In short, the nature of institutional kinds is determined by what a certain group of people collectively accepts. For instance, the kind *money* in the UK depends on the fact that a relevant group of people in the UK accepts the fact that certain things such as bills and electronic currency count as mediums for exchange value (this is overly simplistic, but it is enough for now).

institutional kinds can be Concrete (like driver licences) or Abstract (like laws), or sometimes both (like money).

Thomasson distinguishes two ontological principles to account for these two kinds of institutional kinds. Let us take consider the case of Concrete institutional kinds first:

- Dependence Principle (C): “Necessarily, for all x, x is K if and only if there is a set C of conditions such that it is collectively accepted that (for all y, if meets all conditions in C, then y is K), and x meets all conditions in C” (p. 587).

In other words, the idea is that institutional kinds depend on the fact that we collectively accept a set of required conditions that the kind has to meet, and the kind actually meets those conditions.

Now, the point is to show that epistemological and semantic constraints for these kinds are not compatible with those of natural kinds for the realist.

At a first sight, it is clear that, if accept the C-Dependence Principle, the Error Principle of the realist fails to obtain: if we collectively accept that meeting all conditions of C is sufficient for something to be K, it cannot be the case that meeting all conditions of C is not sufficient for something to be K. Also, since we have to collectively accept the conditions C, it cannot be the case that we are universally ignorant about these conditions (we may be aware of some of them only). In a formal fashion, Thomasson states a new epistemological principle for Concrete institutional Entities:

- Epistemological Principle (C): “Necessarily, for all sets of conditions C, if we collectively accept that (for all y, if y meets all conditions in C, then Ky), then for all x, (if x meets all conditions in C, then Kx) (p. 588).

This means, as before, that our acceptance of the sufficiency of meeting all the conditions in C for being K is sufficient for this to be the case.

Accordingly, also the Causal Theory of Reference fails to obtain: since our beliefs and concepts play a role in determining the nature of Concrete institutional kinds, it is not the case that pre-existing boundaries determine the extension of our terms: there are no pre-existing boundaries.

Similar considerations apply to the case of Abstract institutional Entities. Let us take a quick look at the mechanism in this case.

According to Thomasson, Abstract institutional Entities respond to a new ontological principle:

- Dependence Principle (A): "Necessarily, there is some x that is K, if and only if there is some set of conditions C such that it is collectively accepted that (if all conditions in C are fulfilled, there is something that is K) and all conditions of C are fulfilled" (p. 587).

This principle differs from its Concrete counterpart only because it is not about individual concrete entities acquiring some special function but about the creation of non-physical entities having a specific function.

As before, the Error Principle of the realist is bound to fail: collective acceptance implies that we cannot collectively be wrong. Thomasson defines a new epistemological principle for Abstract institutional Entities. I will not report it here since it is almost the same of its Concrete counterpart. The point is that for the same reasons as above, Causal Theories of Reference fail also with respect to Abstract institutional Entities.

In conclusion, Thomasson showed that social constructivism and social naturalism are incompatible. This is, for our purposes, very relevant since Reality Engineering is committed to social constructivism and rejects social naturalism.

5.1.4 Thomasson on the epistemology of the Constitution Account

The third problem of Searle's ontology, according to Thomasson, is the "self-referentiality" of social concepts. According to Searle, social entities are self-referential in the sense that being thought (or believed, used, and considered) to be F is necessary for anything to be F.

This, Thomasson says, implies that social facts created by collective intentionality are "epistemically transparent" (Thomasson, 2003 (a), p. 274). Something is epistemically transparent if its existent implies that it is believed to exist by the relevant group of people whose collective acceptance guarantees its existence.

Therefore, within a community, there are no social kinds that the members of the community are ignorant about. For something to be of a certain social kind, it is required that everyone in the community accepts the constitutive rules for anything to count as a member of such kind.

However, Thomasson observes, if everyone already knew everything about all the social kinds of his community, then the social Sciences would have nothing to study and nothing to add to these previous beliefs.

She goes on calling kinds which exist even if no one believes them to exist "Epistemically Opaque" (Thomasson, 2003 (a), p. 275) and kinds which exist even if no one has any beliefs about what the relevant K is "Conceptually Opaque" (ibid).

Then, using a similar strategy to the one displayed with respect to the Priority Problem, she shows some examples of social kinds which are not epistemically transparent, such as *recession* and *racism*. Racism can exist in a society even if no one believes that racism exist and even if no one has in mind the concept of being racist.

Therefore, she argues that "not all social entities arise through human agreement *about them*; some social entities (such as recessions) may arise as byproducts of our collective beliefs, practices, and existing institutions" (Thomasson, 2003 (a), p. 278). In other words, epistemically transparent kinds can generate epistemically opaque kinds which nonetheless have the same ontological status.

5.1.5 Thomasson's proposal

Thomasson thinks that analogy with make-believe games can help us understanding the metaphysical structure of social kinds.

Starting off from such an analogy, she distinguishes three different ways in which we can create social kinds.

- The Singular Rule: The most simple make-believe games involve *de re* demands: "they require that we imagine something *of* a particular object" (p. 280). Similarly, the easiest way to create a social entity is to collectively accept that a certain object has a certain social function. The principle can be expressed as follows:

    i. "(Of *a*) We collectively accept: S*a* (where "S" names a social feature)" (Thomasson, 2003 (a), p. 280)

- The Universal Rule: A slight complication is represented by games of make-believe where we are required to imagine that anything of a certain kind, even if it is not actually present or visible, has a certain special feature. Similarly, we can create social entities by imposing rules that say that everything meeting certain conditions has certain social functions. We can express it in the following way

    ii. "For all x, we collectively accept that (if x meets all conditions in C, then Sx)" (Thomasson, 2003 (a), p. 281)

- The Existential Rule: The most complicated instance of games of make-believe involves *de dicto* demands. We are required to imagine that there is something that has certain features. Similarly, we can construct social entities by

collectively accepting that some new entity exists, and it has particular features. We can express the principle this way:

iii. "We collectively accept that (if all conditions C obtain, then there is some x such that Sx)" (Thomasson, 2003 (a), p. 282)

These three rules account for the existence of concrete (i and ii) and abstract (iii) social entities. Therefore, Thomasson has shown how we can create genuine social entities, not just adding new functions to pre-existing objects. Note that iii accounts for epistemically opaque kinds: we are required to have beliefs about conditions and not about the kinds themselves, which we might be ignorant about.

Despite the analogy with make-believe games, social kinds should not be considered as fictional entities. If a relevant group of people G collectively accepts that x is S, then that is sufficient for x to be S. Given the success of the ontological rules displayed above, nothing else is required: the existence of the relevant kind is guaranteed, and its reference is not pretenseful.

5.1.6 Recap of the view and potential problems

Let me recap what I discussed in this section, before I move on to highlighting the good and bad features of the Constitution Account.

First, I presented Searle's view, according to which social kinds are generated via the impositions of social functions on physical objects via the collective acceptance of constitutive rules of the form "a certain x "counts as y in the relevant context C" (Searle, 1995, p. 44).

Then, we discussed Thomasson's criticism about the relevance of physical objects, the relationship between social constructivism and social naturalism, and the epistemic status of social kinds under the Constitution account.

Afterwards, I presented Thomasson's positive proposal, which differentiates between three different patterns of constitutive rules: Singular Rules, Universal Rules, and Existential Rules. Thomasson's proposal severs the connection between social kinds and physical objects and accounts for the existence of social kinds which are Epistemically Opaque.

The good features of the Constitution Account, at least for what concerns Thomasson's version of it, are the rejection of social eliminativism and social naturalism, the strong ontological status of social kinds and the role of social functions. Social eliminativism is rejected because of Thomasson's dismissal of the worry that social kinds are not real. Social naturalism is dismissed by Thomasson's discussion of Dependence and Epistemological principles. Social kinds are granted a strong ontological status because of the constitutive rules. Social functions are relevant because they are what grounds social kinds. The role of social functions is, as we will see, fundamental in all Constructivist accounts. Oddly enough, each position that we will analyse gives a different label for what seems to be the same thing: Thomasson uses the label "function" in 2003 and the label "norm" in 2019; Ásta uses the label "property" and Mallon uses "role". I will not argue for one label against the others, because that all seem to refer to the same kind of phenomena.

There are, however, some problems with the Constitution Account.

First, we have the problem of *acceptance*. It is not clear whether Thomasson and Searle think that acceptance is a mental state that must be possessed by all individuals in a community in order for a certain social kind to exist. Such view would be quite obviously unplausible. Also, it is not clear whether the mental state of acceptance is required *una tantum* when the social kind is created or whether it must be held

constantly. The view that it must be held constantly would be incompatible with the point that acceptance is an active mental state that people have to be aware of. The view that it must be held *una tantum* is weird because even if people changed their mind about the social kind they created, the social kind would remain unchanged.

Related to the problem of acceptance, there is the problem of *collectivity*. It is odd to claim that all members of certain communities must actively accept constitutive rules. One again, that does not seem to reflect the way we ordinarily think about social kinds. This is the biggest problem of the view. This is why we will analyse Ásta's account, which restrict acceptance only to certain members (experts/authorities) within the relevant community. Also, Thomasson herself will move on to a different account where acceptance is substituted by weaker notions, such that not everyone must be aware of the constitution of a social kind.

Lastly, we can see a problem for what concerns social kinds that have people as their members. This is again related to the issues of *acceptance* and *collectivity*. It seems odd to say that people that fall within a certain social category and people that are outside of such category have the same attitudes and the same role in the constitution of the category. A division between the role of external and internal attitudes and roles is required for what concerns social kinds that have people as their members (i.e. social groups). This issue will be dealt with when discussing the Normative Approach.

Chapter 5.2 – The Conferral Account

The Conferral Account is a metaphysical theory of social kinds presented by Ásta in her paper "The social Construction of Human kinds" from 2013[1]. The goal of the author of this paper is to provide a strong metaphysical base to social constructivism, in order to allow for projects of social *debunking*. The author considers herself in line with Sally

Haslanger's account of social kinds, at least for what concerns the project of getting rid of some of them. The Metaphysical part of the account, however, is relevantly similar to the one that we have seen in the Constitution Account, with a couple of adjustments that constitute major improvement over the aforementioned problem of *collectivity*. While presenting the view, I will start off by reporting the reasons and goals that led to its development. Then, I will proceed to an explanation of the view. Next, I will discuss the points of similarity and those of difference between the Conferral Account and the Constitution Account. I will go on discussing Ásta's idea of *debunking* projects, showing how it relates to my idea of Reality Engineering. Lastly, I will provide a recap of the view, highlighting its good and bad features.

5.2.1 Elucidations on a social Constructivist metaphysics

social constructivism (the view that social kinds exist mind-dependently) has often been accused of accounting for social kinds that are not *real* enough. There is an odd prejudice about social constructivism, and many seem to imply that if something is socially constructed, than it is not real. In this dissertation, I have tried to resist this prejudice on many occasions. At the beginning of her paper, Ásta faces the same type of issue.

She does this by separating two claims of social constructivism:

(1) The Epistemic Claim: Our conceptions about a certain social category are influenced by our social practices.

(2) The Metaphysical Claim: The existence and the ways of existing of the social category in question is influenced by our social practices. (explicit paraphrases, in my terms, of Ásta, 2013, p. 716)

According to Ásta, early sympathizers with the social Constructivist framework (such as Hacking (1999) and Boghossian (2006)) have endorsed the Epistemic Claim and rejected the Metaphysical Claim. Once again, the reason is that they feared accusations of providing a *weak* metaphysics. Ásta's goal, similarly to Thomasson's (and mine), is to show that we should commit to the Metaphysical claim and reject such accusations.

The point is that socially constructed entities can have a robust ontological status, as long as we are explicit and rigorous about the analysis of their foundations. Socially constructed social kinds are a fundamental component of the world we live in, and a fundamental component of everyone's life. Therefore, a metaphysics that dismisses them is not one that we should commit to, due to an obvious failure to accommodate entities that are extremely important for theoretical and practical purposes.

5.2.2 Explanation of the Conferral Account

So, with the goal of providing a robust metaphysics for socially constructed kinds, Ásta develops the Conferral Account.

The key notion of the account is the one of "Conferred property" (Ásta, 2013, p.719), which is defined as a property "that something has in virtue of some attitude, action, or state of subjects, or group of subjects" (Ásta, 2013, p. 719).

To clarify the definition, she uses a famous example from the history of philosophy. In Plato's dialogue *Euthyphro*, Socrates and Euthyphro discuss the nature of the property *being pious*. Socrates argues that the gods love certain actions because they are *pious*. Oppositely, Euthyphro argues that an action can be defined *pious* when it is loved by the gods.

Euthyphro's position is the ancient relative of the Conferral Account. The property of *being pious* is attributed (conferred) by someone (the gods) to something (actions) in virtue of some attitudes (love of the action).

So, by starting with this easy example, we can show Ásta's five pillars of the view (Ásta, 2013, p. 720):

(1) *Conferred property*: the property that is being conferred (e.g. being pious).

(2) *Who*: the subject/group of subjects that confer the property (e.g. the gods).

(3) *What*: the relevant attitude that generates the conferred property (e.g. love).

(4) *When*: the context in which the conferred property is generated.

(5) *Grounding property*: what the subject is trying to track, if anything.

5.2.3 Conferral vs Constitution

In order to provide an analysis of Ásta's five pillars, I will use the Constitution Account as a dialectical alternative, in order to show similarities and differences with the Conferral Account and clarify issues about the latter. Let me break down the discussion in five parts, following Ásta's formulation of the five pillars.

(1) Conferred properties vs social Functions

The first pillar of Ásta's view is the one which, although under different labels, is shared by all the views that I am going to present. In fact, Ásta's Conferred properties seem to be the exact same thing as Searle's and Thomasson's social Functions. They

both represent the *extra* non-natural feature that is attributed to something and that makes it distinctive with respect to ordinary natural entities. Indeed, it is easy to see that function-talk can be translated in property-talk with just a little trick: if I add a social function to a certain object, then I also add to the object the property of having such social function.

(2) Restricted vs Unrestricted groups of subjects

A fundamental point of difference is represented by Ásta's answer to the *who* question. Ásta believes that attitudes of a restricted group of subjects within a community are enough to confer certain properties. Oppositely, we have seen that one of the most prominent problems of the Constitution Account is the problem of *collectivity*: the view claims (implausibly) that the active mental state of acceptance of all members of the relevant community is required for the constitution of a social kind. Ásta avoids the problem by implying that, depending on the context, different restricted groups of subjects are in a position such that their attitudes count for the conferral of a certain property. Under to this light, we see that Ásta's account introduces (implicitly) a figure that we will later call *authority*: an individual which occupies a special place in a community, such that she gets to participate in the process of creation of social kinds.

(3) Attitudes

The role of attitudes in the Constitution Account is not as explicitly relevant as in the Conferral Account. However, we can read the Constitution Account as saying that the attitude of *acceptance* is required for the creation of a social kind. This opens the floor for an interesting comparison. According to Ásta, the role of subjects under the Constitution Account is merely an epistemic one: once the constitutive role is in place, a subject can only "discern what the [social] fact already is" (Ásta, 2013, p. 720). On the

contrary, under the Conferral Account the subject's role is a metaphysical one: his/her attitudes are the metaphysical grounding of the social fact in question. I have to say that Ásta's reading of the Constitution Account is a bit confusing, since it seems that subjects might have an epistemic role once the rule is set but they definitely have a metaphysical role when setting the rule. But this discussion does indeed shed light on the issue of *acceptance*. We see that, according to Ásta, the attitude that is relevant for the conferral of a certain property must be present in *every instance* that we wish to assert the presence of such property. Oppositely, Ásta thinks that under the Constitution Account the relevant attitude (acceptance) is required only *once*, and she takes this to be a problem.

(4) What counts as a Context?

The issue of Context is treated quite similarly by the two accounts. Ásta is more explicit in her examples about the variety of possible contexts in which social entities have a role. Under her framework, a baseball game is a context in which specific social entities are at work. The Constitution Account does not really deal with this issue, but the two accounts are probably compatible under this light.

(5) Is Grounding Necessary?

For what concerns the role of physical objects, Ásta seems to be in accordance with Thomasson and in disagreement with Searle. As we have seen, one of Thomasson's main contributions to Searle's theory was the cutting of the connection between social entities and physical objects. By saying that subjects do not need to be tracking any grounding property when conferring a social property, Ásta is claiming the same thing. This move is helpful in two ways: first, it helps accounting for social kinds that do not make sense in terms of relation to physical objects ("the U.S. Constitution, General

Motors, or the Calvinist doctrine" (Thomasson, 2003 (a), p. 273); secondly, it allows for social kinds which are by-products of other social kinds, and therefore not grounded on the physical world. About the first point of separation of social and physical entities, Ásta has an interesting contribution:

> On a conferral account, however, whatever there is in the object plays no causal role, only an epistemic one. There can be something in the object that the subject is tracking, but it plays no causal role. (Ásta, 2013, p. 721)

In other words, it is not always the case that there is something in the physical objects that we try to account for when conferring a social property. And even if there was such feature, its role would be just of being the content of an attitude (hence, an epistemic role), and the of causing or grounding such attitude.

About social kinds that are by-products of other social kinds, we have seen Thomasson's discussion of Epistemically Opaque kinds as kinds that appear as by-products of other more basic kinds (like *depression* which is grounded on the kinds *money* and *market*). Ásta makes a similar claim by saying that social kinds are disposed in hierarchic fashion. Let us see how she puts it:

> Apart from the legal property *being a legal refugee* that comes with legal privileges and burdens, there is also another conferred property *being assumed to be a legal refugee*, which comes with its own social constraints and enablements and that in the conferral of this latter property the property *being a legal refugee* is being tracked. (Ásta, 2013, p. 728)

In other words, the property of *being a legal refugee*, for Ásta, is grounded on the property of *being assumed to be a legal refugee*. Similarly, one could argue that *gender* is a social kind which is grounded on *sex* which is also a social kind, and that only the latter actually attempts to track something in the physical world (similar discussion in Ásta, 2013, pp. 725-726).

The common point between Ásta and Thomasson is that some social kinds produce other social kinds as by-products, but those latter ones should not be considering any less or more real than those former ones.

5.2.4 Debunking and Reality Engineering

Ásta developed the Conferral Account with a clear goal in mind: providing a metaphysical background to *debunking* projects in social metaphysics. A debunking project is a project that aims to show that a certain social category is not natural and inevitable. Therefore, if we see that the category has bad effects in society, we can get rid of it.

Let us be careful about not falling into a confusion here: the idea that a social category should be *eliminated* (other word for *debunked*) does not mean that we commit to social eliminativism. Social eliminativism claims that social categories do not exist in the first place. Debunking projects of social constructivism say that social categories exist, that they exist mind-dependently, that they are defective, and that we should get rid of them. Along with this point, I would like to remark once again that the noble background of social eliminativism of rejecting the defectiveness of some social categories should not lead us to believe that those social categories do not exist in the first place.

Ásta, similarly to Haslanger, argues that *gender* is a socially constructed, and defective, social kind. For this reason, we are entitled to advocate for an elimination of such social kind, so that society can finally get rid of such an oppressive and defective category. The aim is the creation a gender-neutral society, where no one is discriminated for her/his sexual orientation or gender identification.

We can see that the reasons and goals of this *debunking* attitude towards certain social kinds unite Ásta's view and my account of Reality Engineering. In Reality Engineering we observe that certain social kinds are defective, and we try to modify them, with a metaphysical background that tells us that those categories are not natural and not inevitable. Reality Engineering is, however, much more than a mere *debunking* project. As we have discussed above, Reality Engineering via *elimination* is just one of four ways in which categorial modifications can occur, alongside with *revision*, *replacement*, and *introduction*. Ásta's method of elimination is therefore a subset of Reality Engineering. The reason why *elimination* is not always the best option is that some social Categories are good under some respects, and not only defective. Their elimination would leave a gap in our ontology and in our conceptual space that we might want to fill with some renewed and ameliorated kind. The elimination of all defective social kinds would indeed make our social ontology poor of theoretical and practical interest. What we want is a project that eliminates defectiveness but still fulfils the function of helping us navigate our social lives and improve them.

5.2.5 Recap of the view and potential problems

First, we saw that Ásta provides a compelling argument to reject social naturalism. She does so by defending the Metaphysical Claim of social constructivism that says that the existence and the ways of existing of social categories are influenced by our social practices.

Then, I showed the five pillars of Ásta's Conferral Account: *conferred properties*, *groups of subjects*, *attitudes*, *contexts*, and *grounding properties*. I then explained these points by comparing them to the corresponding features of the Constitution Account.

What emerged as particularly compelling in Ásta's account is the avoidance of the problem of *collectivity* (that weakens the Constitution Account) and the clarity with which Ásta opened the path for debunking projects, which is relevant as a background for Reality Engineering.

There are a couple of points that are less compelling about this account:

First, as we discussed in the previous section, one might be puzzled about the idea of *debunking*. We have seen, with Thomasson, that social kinds are fundamental for us to navigate our social lives. Getting rid of them without providing an alternative seems to leave too many gaps open, creating theoretical and practical issues.

Secondly, one could object that the constant role of authorities in the process of conferring social properties seems implausible. Surely this view makes sense in contexts like a baseball game, that Ásta uses as an example. In such context, the constant presence of a referee is essential in order to determine which action count, for instance, as fouls or points. But this does not translate well in the case of more complex contexts. Let us use the other example by Ásta to show this. Ásta talked about the property of *being a legal refugee* as grounded in the primitive conferred property of *being assumed to be a legal refugee*. However, it does not seem that for every refugee there is some authority that has the attitude of assuming that he is a refugee. Therefore, under the Conferral Account, she would not count as a refugee. In this respect, the Constitution Account worked better: once the constitutive rule is set, then individuals can fall into a social category even without anyone having specific attitudes about them.

Lastly, one might see some issues with Ásta's property-talk. Ásta reduces membership to a certain social kind to the possession of a single conferred property. Although this has the advantage of easily individuate members of a certain social kinds (all and only

those individuals that possess such single conferred property), it also has the risk of leaving out individuals that we would like to include in the kind but lack that specific property. As we are going to see in the next section, social kinds are usually identified by multiple properties that individuals possess. Not all members of the kind possess all those properties, but they count, nonetheless, as members of the kind. Ásta's account does not allow for these individuals to be included in the relevant social kinds, and this is undesirable.

Chapter 5.3 – The Entrenchment Account

The Entrenchment Account is a view that comes from the work of Ron Mallon and Jonathan Tsou. The label that I chose to name this view is not found in neither of these two author's publications, it is just my way of naming the view.

The discussion of this account has sparked from Mallon's book "The Construction of Human kinds", from 2016. But for the purposes of this dissertation, I will mainly use Tsou's paper "social Construction, HPC kinds, and the Projectability of Human Categories", from 2020. As in the case of the Searle-Thomasson Constitution Account, I decided to refer mainly to a later interpretation of the view, since it provides interesting insight.

I will present the Entrenchment Account as follows. First, I will locate this account in the social Constructivist framework, discussing the relationship of this position and social naturalism (which, I think, caused some confusion in the authors too). Secondly, I will analyse the metaphysical aspect of the view, which is mainly due to Mallon. Then, I will discuss Tsou's contribution on the epistemological aspect of the position. In conclusion, I will recap the view and highlight the good and bad features of it. I will argue that despite interesting metaphysical and epistemological premises, Tsou draws a

rather misguided and misleading conclusion. By getting rid of these conclusions, we will be able to appreciate the compelling metaphysical picture that the account provides.

5.3.1 Elucidations on a Social Constructivist metaphysics (again)

The Entrenchment Account in Mallon's original version was developed in order to show that social constructivism is compatible with a realist metaphysics. Mallon's goal was to show that social kinds are "*real and causally predictive*" (Tsou, 2020, p. 116), so that they allow for projectable inferences.

Under this light, it seems that Mallon wants to unite social constructivism and social naturalism. This, however, would be a mistaken interpretation. To be fair, it is a mistake caused by a confusion in the author's view. What Mallon is doing when arguing that social kinds allow for projectable inferences is to provide a robust epistemological status (erroneously labelled *natural*) to a social Constructivist metaphysics. Social naturalism as a metaphysical framework has nothing to do with these arguments. Nowhere in Mallon or Tsou's work there is a remark about social kinds existing mind-independently. Therefore, Mallon's account is not conflating opposite metaphysical frameworks, as both Mallon himself and Tsou seem to imply.

As we will see in a moment, Mallon's metaphysics is a typical instance of the social Constructivist framework. The idea that social kinds exist mind-dependently and that they have objective epistemic standard is nothing particularly revolutionary. We saw that also under Searle's version of the Constitution account social kinds are metaphysically subjective and epistemically objective. So, Mallon's goal is not that different from Searle's. The interesting part of the account is the completely different metaphysics that Mallon provides.

Then, it can be debated whether a social Constructivist metaphysics should have robust or weak epistemic statuses. For instance, Mallon thinks that such a metaphysics is in fact supported by a strong epistemology and Tsou thinks the opposite. None of these arguments, however, has anything to do with social naturalism as a metaphysics.

Once again, I would like to remark that we should not fear accusations of accounting for social kinds that are *not real enough* under social Constructivist views. Social entities which are socially constructed have a perfectly legitimate ontological status, as we showed in multiple occasions.

5.3.2 Social Kinds under the Entrenchment Account

The purely metaphysical aspect of the Entrenchment Account is much more complex than the ones that we have seen with the Constitution and Conferral Account. Under those previous frameworks, social kinds are generated by the stipulation and acceptance of a certain rule: in the first case, the rule counts as an attribution of a social function to a physical entity (Searle); in the second case the rule counts as an attribution of a conferred property. In both cases, the metaphysical grounding of social kinds is straightforward. Mallon's metaphysics differs deeply from these two views.

According to Mallon, the key notion that we have to keep in mind when thinking about the construction of social kinds is the one of "social Role".

In Mallon's account, social Roles are individuated by representations like "attitudes, theories, narratives, concepts, models, pictures, norms, rules, utterances" (Mallon, 2016, p. 6) and categories are defined as properties (intensionally) or classes of individuals (extensionally).

He points out two necessary existence conditions for social roles. A social role exists iff:

1. “There are representations that pick out a category of persons[2] and a set of beliefs and evaluations associated with that term” (Tsou, 2020, p. 117).
2. “Many of the beliefs and evaluations associated with the role are common knowledge within society” (Tsou, 2020, pp. 117-118).

In other words, a social Role exists if members of a community commonly share some beliefs and attitudes towards a certain group of persons within such community.

Let me highlight that this account of social Roles is parallel (with some legitimate differences) with Searle and Thomasson’s account of social Functions and Ásta’s account of Conferred properties. Given what we have seen with the explanations of those accounts, we might expect that the existence of a social Role will count as sufficient for the existence of a social kind, in the same way in which Conferred properties and attributions of social Functions counted as sufficient in the other accounts.

Mallon, however, thinks that the mere existence of a social Role is not sufficient to construct a social kind. He thinks that a social Role *may* lead to the construction of a social kind if it surpasses a phase that Tsou later conveniently labels *Entrenchment*. Mallon’s idea is that social Roles must *pass the test of the world* before they can turn to legitimate social kinds. Such test consists in the contrast of social Roles with psychological and environmental mechanisms. Social Roles could be transformed (or dismissed) by “Behavioral Influences” (Mallon, 2016, p. 69) and/or “Environmental Construction” (Mallon, 2016, p. 82). By behavioral influence, Mallon means that social Roles, in terms of people’s attitudes towards other people, can have impact on those people’s behaviours. In turn, these behaviours can cause transformations of the attitudes that identified the social Role in the first place. Therefore, due to different behaviours

social Roles can evolve in different ways[3]. Similarly, social Roles might cause transformations in the environment (understood as cultural and institutional). In turn, the environment might end up accepting and including the social Role (and thereby making it a social kind), or rejecting the social Role, or transforming it.

So, if social Roles surpass Behavioral and Environmental obstacles, then they turn into legitimate social kinds. These means, in easy terms, that people (or objects) that were only *thought to be* part of a category during the social Role stage, *actually become* part of a category.

Another interesting feature of the account is that the emerging social kinds are considered to be Homeostatic property Cluster kinds (HPC kinds from now on). The concept of a HPC kind was first introduced by Boyd, in his book from 1999 which dealt with natural kinds. HPC kinds are individuated by multiple properties (clustered properties), which have the following features:

Co-Occurrence: "[These properties] are contingently clustered in nature in the sense that they co-occur in an important number of cases." (Boyd, 1999, p. 143f)

Homeostasis: "Either the presence of some of [these] properties tends […] to favor the presence of the others, or there are underlying mechanisms or processes that tend to maintain the presence of [these] properties, or both." (ibid.)

To explain, HPC kinds are individuated by properties that tend to be possessed at the same time by a certain individual and this tendency is motivated by underlying mechanisms that somehow link them together.

The point of interest about HPC kinds under Mallon's account is the application of Boyd's theory to the social kinds debate. The convenient idea is that social kinds are not individuated by a single property. Members of the same kind can have different

properties from one another, and as long as these properties are all part of the relevant cluster, they still count as members of the kind. There is not a single essential property that must be possessed in order to qualify as a member of a kind.

We have seen that, under the Conferral Account, the possession of one single Conferred property was necessary and sufficient for kind-membership, and that was problematic. Under this respect, Mallon's view takes a step in the right direction.

5.3.3 The Epistemology of Social Kinds

social kinds under the Entrenchment Account are constructed by the entrenchment of social Roles with psychological and environmental mechanisms. This entrenchment produces HPC kinds.

Mallon's goal, when developing this metaphysics, was to show that social kinds are able to generate projectable inferences and that they are real and causally predictive kinds. In easy words, Mallon wants to account for social kinds with are useful for explanations and predictions.

Tsou, in contrast, aims to show that social kinds cannot be as useful for explanations and predictions as much as natural kinds are. According to Tsou, this argument would imply the downplaying of the role of social kinds within the realm of the social Sciences.

Tsou's argument is based on a distinction between two different kinds of inferences: Robust Projectable Inferences and Transitory Projectable Inferences.

According to Tsou, an inference is a Robust Projectable Inference (RPI from now on) iff it is *ampliative*, *contextually stable*, and *temporally stable*.

An inference is a Transitory Projectable Inference (TPI from now on) iff it is either *temporally unstable* or *contextually unstable*. TPI can be both *ampliative* and *non-ampliative*.

To clarify, let me provide some further definitions of the technical terms used by Tsou.

An inference is *ampliative* iff it is not just a matter of definition. For instance "widows are women" (Tsou, 2020, p. 122) is just a matter of definition, since the definition of *widow* includes the fact that widows are women. Oppositely, 'cheetahs are fast animals' (my example) is not just a matter of definition, but it is an inference that follows from observation. The latter is an instance of an ampliative inference. The former is an instance of a non-ampliative inference.

An inference is *contextually stable* if its soundness is not affected by the social context in which it is made. Social and historical factors may make an inference sound within a certain community and not in others.

An inference is *temporally stable* if its soundness is not affected by the historical moment in which it is performed. For instance, the inference 'water boils at 100 degrees' is sound independently of the time of its performance; while the inference 'having slaves is unacceptable' is sound now but was not sound until a few centuries ago (in some places, due to contextual instability, a few decades ago).

Tsou's point is that social kinds can only produce TPI inferences, while natural kinds produce RPI inferences. This follows from the, quite straightforward, analysis of Mallon's metaphysics. Observing the metaphysical foundations of the Entrenchment account, we see that social kinds are grounded on social Roles and psychological and environmental mechanisms. Even without too much dwelling, it is easy to see that both social Roles and those mechanisms are temporally and contextually unstable. Social

Roles are defined as sets of beliefs and attitudes. They change over time and they differ from culture/or community to culture/or community. Same for psychological mechanisms and environmental features (obviously contextually unstable).

So, Tsou's argument goes through: social kinds (under the Entrenchment Account) can only produce Transitory Projectable Inferences.

Tsou's controversial conclusion is, however, that social Sciences should look for social kinds which are grounded only on intrinsic, natural, properties, since they are the only ones that can produce Robust Projectable Inferences. Hence, social Sciences, according to Tsou, should take a naturalistic turn.

5.3.4 Recap of the view: Good Premises, Edgy Conclusion

Let me recap the view and discuss some problems with the conclusions that Tsou draws from it.

The Entrenchment Account is composed by two parts: a metaphysical analysis of the nature of social kinds, and an epistemological analysis of what kinds if inferences can be produced by social kinds.

About the metaphysical issue, we have analysed Mallon's theory that social kinds emerge as HPC kinds via the entrenchment between social Roles and psychological and environmental mechanisms. We defined social Roles as sets of beliefs and attitudes (broadly) which are common knowledge in society. Also, we defined HPC kinds as kinds that are individuated by clusters of properties that are glued by co-occurrence and homeostasis.

About the epistemic side, we have discussed Tsou's notions of Robust Projectable Inferences and Transitory Projectable Inferences. We have seen that social kinds can

only produce Transitory Projectable Inferences, i.e. inferences that are contextually and/or temporally unstable.

Both the metaphysical and the epistemic views are compelling and full of merits. Just to mention a few of these merits, we have seen how the Entrenchment Account fixes the problem of Ásta's property-talk. Kinds are no more individuated by single, isolated properties. Rather, members of a kind are allowed to have different features from one another, but still fall into the same category. Also, Mallon's idea of "common knowledge" gives us some insight into the problem of *collectivity* that we have seen in the Constitution Account: common knowledge is a much more passive notion compared to collective acceptance. This means that we might have common knowledge of something (as a community) even if not everyone has active mental states and attitudes towards that thing. This is convenient because Searle and Thomasson's notion of collectivity seemed to imply that every member of a community must have attitudes towards the rules that make something a social entity, and we discussed the implausibility of such claim.

Another merit of the view is the clear explanation of the epistemology of social kinds and their role in the production of Transitory Projectable Inferences. As Tsou himself says: "For philosophers interested in social change, it should be regarded as a good thing that these sorts of projectable inferences are neither necessary nor stable" (Tsou, 2020, p. 124).

Within this research project, we are indeed philosophers interested in social change, and Tsou is right in pointing out that his epistemology counts as good news for us.

However, Tsou claims that, due to such epistemology, social Sciences should take the naturalistic turn of focusing on social kinds which are grounded on intrinsic natural

properties. I will not enter a discussion about the role of social Sciences here, but I think that we should be careful about the underlying metaphysical claim that Tsou is making. Once again, Tsou, by saying that naturalistic social kinds are the only ones good enough to be the focus of social Sciences, seems to claim that such social kinds are somehow *more important* or *more robust* than the others.

I have three remarks against Tsou's implicit claim. First, it is not at all clear how these naturalistic social kinds can be accounted for under Mallon's metaphysics (which is the framework under which Tsou is making these claims). I cannot understand how social Roles could pick out natural properties. Secondly, and more importantly, we get the impression, once again, that a Constructivist metaphysics should be considered inferior to a naturalist metaphysics. I think I provided enough arguments to debunk this prejudice. Third, Tsou hugely underestimates the importance of Transitory Projectable Inferences. As members of communities, located in a certain context and a certain time, we are extremely interested by the inferences that are produced by the social kinds that exist here and now. It is not a problem that these inferences can cease to be sound in other contexts/times, as long as we are aware of it. Social Sciences can be hugely helpful for us in the navigation of our social lives, even if their claims are only contingently true.

Chapter 5.4 – The Normative Approach

The Normative Approach is mainly due to Amie Thomasson, who developed the view in a paper from 2019 called "The ontology of social groups". Some interesting insight is provided by Katherine Ritchie, but we could also read some of Haslanger's work as a base for the view. Furthermore, the account is also discussed in a paper that I co-authored (Kaspers, LiBrizzi, Calosi, Kobe, *under review*).

In presenting the account, I will proceed as follows. I will start by discussing how the view fits with respect to the framework of social constructivism, discussing Thomasson's arguments about the matter. I will also discuss Thomasson's anti-metaphysical approach[4], since she thinks that the account is not providing metaphysical guidelines on the nature of social kinds. Then, I will proceed to an explanation of the view, translating Thomasson's claims in metaphysical terms. Later, I will discuss the notion of "external norm", which is key for the account. Next, I will proceed with an argument for the sufficiency of the presence of external norms for the existence of social kinds. I will wrap up the discussion by recapping the view and highlighting some debatable features of it.

5.4.1 Elucidations on a Social Constructivist metaphysics (one last time)

The relationship between Thomasson's Normative Approach and social constructivism as a metaphysical framework is controversial. Such controversy is due to Thomasson's anti-metaphysical approach. In 2015 Thomasson argued for an *easy* approach to ontology, where existence questions are answered straightforwardly just by conceptual analysis and empirical observation (Thomasson, 2015). Also in this paper from 2019, she maintains the same approach by claiming that the answer to existence questions about social kinds is "an easy and obvious 'yes'" (Thomasson, 2019, p. 4829).

This counts as good news for a social Constructivist metaphysics, since it is a claim that shuts down the worry brought up by social eliminativism, according to which social kinds do not exist at all.

The puzzling feature of Thomasson's view is that, despite her confidence about the positive outcome of ontological questions, she is not attempting to provide a metaphysics for social kinds. Rather, her goal is explicitly of setting aside metaphysical

questions, in favour on epistemic and practical ones: instead of "What are social groups?" we should ask "What do we want the concept of a social group for?" (Thomasson 2019, 4836).

Thomasson proceeds to an analysis of the function of social kinds (which I discuss in the next paragraph), claiming that she is not providing necessary and sufficient conditions for the existence of such kinds. Under her view, an analysis of function in terms of norms that are connected with social kinds has only epistemic and practical significance, and not a metaphysical one.

In the paper that I co-authored, we reject this anti-metaphysical approach. The point is that Thomasson's account does not succeed in its goal to avoid metaphysical questions. Thomasson's epistemic claims can be viewed as metaphysical ones as well. In the paper, we stated that:

> Thomasson's ontology is not as "easy" as she would like it to be, and that some deeper metaphysical work is required. Indeed, we think that an ontological interpretation of Thomasson's normativity-based criterion has some very interesting metaphysical implications that one could only appreciate if one rejects Thomasson's metaontology. (Kaspers, LiBrizzi, Calosi, Kobe, *under review*, p. 3)

For the purposes of this research, I will maintain the same approach, and I will treat Thomasson's views as if they were an analysis of the metaphysics of social kinds.

From the standpoint of a social Constructivist framework, however, we can extrapolate a valuable lesson from Thomasson's position. The point is that metaphysical questions about social kinds, by themselves, are of little interest. The goal is always to link such metaphysical discussions to practical ones, where we discuss the role, function, desirability, and the potential to change of social kinds. This dissertation is profoundly inspired by Thomasson's approach in this respect: my aim is to analyse deeply the metaphysics of social kinds in order to understand how we can engineer them.

5.4.2 Social groups under the Normative Approach

Before describing Thomasson's approach and drawing metaphysical conclusions from it, I need to address a preliminary point.

Thomasson's view is intended to account for social groups, and not social kinds. The difference, contrary to the difference between social kinds and Human kinds that we have seen before, is not dismissible. Thomasson's position, at least in its original version, is built to account for categories which have human individuals as their members (i.e. social groups). I will make the case that the revised version of Thomasson's account found in the paper that I co-authored can be extended to account for social kinds, and not only social groups. But let us keep in mind, for now, that the scope is just social groups.

As we started to see above, Thomasson's position focuses on the issue of *function*: what do we need social groups for? Here is how Thomasson puts it:

> Many of our most central social group concepts have *significance* for us because they function to normatively structure our lives together: marking how we are to trat others and how we are to behave in a variety of contexts and towards a variety of people. (Thomasson, 2019, p. 4838)

In other words, the relevance of social groups (and their respective concepts) is that they give us guidelines for behaviour: they set the standard for how we have to treat other people and how we should behave in different situations.

This analysis fully captures the reason why it is of fundamental importance to recognize the existence of social groups (and social kinds). Here is a further point by Thomasson:

> An individual who lacked all social group concepts would, quite simply, have no idea how she was to act, what was expected of her, or how to understand the reactions of others.
> (Thomasson, 2019, p.4838)

To give a slogan to this analysis, we can say that: practical significance implies ontological robustness.

Thomasson proceeds to formalise the idea of a *normative structure* by differentiating between three types of norms: internal, structuring, and external. These norms characterize social groups in the sense that there is a general consensus about their validity, and everyone behaves in accordance with them (more on this in the next section).

Internal norms are "norms regarding how members of the group are to behave, regard themselves and other group members" (Thomasson, 2019, p. 4838). These norms give guidelines to people who are *inside* the social groups about which attitudes and behaviours are appropriate to have *qua* being members of the group. Such attitudes and behaviours are specific to single groups, and the only reason that one must have in order to follow them is to self-identify as a member of the group. To give an example, *as a member* of a basketball team, I stand by the norm of showing up to practice every time. For me, there is no other reason for showing up at practice if not for my membership to the team.

Structuring norms are "norms that place different members of the group at different nodes, with different norms regarding those who occupy different nodes" (ibid.). In cheap terms, structuring norms tells members of groups whether there is a hierarchy of some sort built into the group, and, if so, what attitudes and behaviours a member of one of the layers of this hierarchy should have with respect to members of other layers. As instance, as a student who is *member* of the social group that is the University, I am

aware of how I am expected to behave with respect to my supervisors and other faculty members.

External norms are "norms regarding how members of that group are to be treated, regarded, behaved towards by those who are not members of the group (ibid., p. 4839)". An easy example is the one of the social group of the *elderly*. We all (mostly) stand by the principle that we should treat the elderly with respect. This is exactly what Thomasson means when talking about external norms: they are sets of behaviours and attitudes that people are expected to have towards members of social groups.

In order to translate this account to metaphysical terms, we must get deeper in the analysis of the nature of external norms and their role, since I will make the case that external norms are the key notion of the account, and that their presence is sufficient for the existence of a social group.

5.4.3 What are external norms?

Thomasson has a deflationary attitude towards the metaphysics of social groups. According to her, social groups obviously exist, and we can see that just by conceptual analysis and practical observations. Also, she says that social groups are individuated by various norms.

Even prior to deeper analysis, it is easy to see that a non-deflationary social Constructivist could say, following these guidelines, that social groups just *are* norms, or at least some kind of reification of norms. But in order to justify this claim we have to look further into the nature of norms, so that we can understand their metaphysical role.

For this reason, I will now discuss the notion of *external norms*. I only focus on external norms because, at it will emerge later on, I think that external norms are sufficient for the Construction of social groups.

There are various questions concerning external norms: one might ask about their origin; about how spread they are (or should be) and about how diversified they can be.

First, let us recall that external norms are "norms regarding how members of that group are to be treated, regarded, behaved towards by those who are not members of the group (Thomasson, 2019, p. 4839)".

About the origin of external norms, we can say that they do not differ much from Mallon's social Roles. Different norms can have different origins: some might come to us by cultural heritage (like the norm of treating the elderly with respect), some might be generated by moral principles widely accepted in our communities (like the norm of not discriminating people for the colour of their skin), some might come up due to exceptional and new circumstances (like norms about the respect of the environment are generated by the current climate crisis), and so on.

The relevant point is that whatever their origin is, external norms must be spread in society in order to count as legitimate norms. This is what Mallon had in mind when he said that social Roles must be *common knowledge*. However, one big question still arises: how *common* should this knowledge be?

The question can be read in two ways, first we can read it as asking how many people should be aware of it in order to count as sufficiently spread. Secondly, we can read it as asking whether everyone that is aware of these norms has in mind the same thing.

About the awareness question, one good way to put it is that not everyone must be actively aware of it. This would lead us back to the problem of *acceptance* typical of the

Constitution Account that we have seen above. The most inclusive way to account for awareness is to put the constraint that everyone has to respect the norm (either consciously or unconsciously) or consciously refute it. So, it could very well be the case that no one ever told me to treat the elderly with respect, but I still do for separate reasons. In such scenario, I am unconsciously respecting the norm. Oppositely, a requirement for legitimate external norms is that whoever wants to violate them must be conscious about them. If someone can violate the norm without being aware of the fact that he is violating it, and thereby not knowing that she could encounter bad reactions by other members of the relevant community, that would mean that the norm is not spread enough. If someone could violate the norm without actually encountering any reaction from other members of the community, then the norm is not spread enough.

There is no doubt that one could object to these constraints for external norms and generate borderline cases. The point is not that this view is the most theoretically adequate, but that it is the most practically inclusive.

About the possibility of people having different versions of the norm in mind, a couple of considerations are important. First, external norms are *contextual*. This means that different communities might share different norms. Therefore, it should not surprise that members of different communities can have very different notions of the norm of treating the elderly with respect if they come from different communities in which, for instance, who counts as elder and what counts as respect differ. The relevant point is that *within* the relevant communities external norms should be shared. Also, it would not be a problem if different people within the community had different notions of the norm in mind, as long as they all behave in similar ways. It is to be expected that we would get diverse answers if we asked people to formulate their version of the norm.

But as long as everyone behaves in accordance with the most general version of the norm, there is no problem.

More on these aspects will be discussed in section 6, when I will provide two patterns of social Constructivist metaphysics.

5.4.4 The Sufficiency of external norms

Now that we have elucidated what external norms are, we can provide an argument for their sufficiency in the Construction of social groups.

In the paper that I co-authored, me and my colleagues made the case that, following Thomasson's account, there are good reasons to conclude that The Dead should count as a social group. Following that claim, it was easy to argue for the sufficiency of external norms for the existence of social groups, since The Dead do not share any internal norms (*qua* being dead).

I believe, however, that we can reach the same conclusions by starting with different premises as well: we do not need instances of social groups that do not have internal norms in order to show that external norms are sufficient.

As a starting point, we should observe that almost all the social groups which are at work in our communities are such that the members share internal norms, and the non-members share external norms. Also, it is clear that internal norms and external norms can influence one another. The way people behave with respect to members of a group might influence the way such members behave and viceversa.

However, we can rule out the possibility of internal norms being sufficient for the existence of social groups. Consider the following example:

Pick some random property, like having a pinky toe that is *x*cm long. Imagine that there are precisely 10 people in the world who have this specific length of a pinky toe, and that they all live in different parts of the world. Each believes that there will be some number of people other than themselves that share this

particular property, but they have never met each other and perhaps they never will. By some miracle, each member of this collection comes up with similar sets of norms regarding how they, the special pinky length people, ought to behave and each of them follow these norms. (Kaspers, LiBrizzi, Calosi, Kobe, *under review*, p. 11)

Now, if internal norms were sufficient for an assemblage of people to count as a social group, then those people would count as a social group. However, that would be rather puzzling. Let us also recall that Thomasson's Normative Approach was motivated by the need of discerning legitimate social groups from arbitrary assemblages. Allowing internal norms to be sufficient would cause a variety of arbitrary assemblages to count as social groups, and this is problematic. Even further:

Of course, these individuals could try to create a genuine group out of this shared arbitrary property, but this would require the presence of some norms that are not internal to the individuals, such as structuring norms (which are internal to the group but are external to the individual members) or external norms. They could do this by getting in touch with potential members of the group and creating a network or by spreading the word to force people outside of the group to recognize the group's existence. These strategies further corroborate the idea that internal norms are not sufficient by themselves. (Kaspers, LiBrizzi, Calosi, Kobe, *under review*, p. 11)

So, there are good reasons to believe that internal norms are not sufficient for the existence of social groups. Let us turn our attention to the argument about the reasons why external norms are sufficient.

It is difficult to find actual instances of social groups which have external norms but not internal norms. This follows from the fact that the way people behave with respect to members of a group influences the way such members behave themselves (and viceversa). The point, however, is that widely spread external norms can individuate social groups even though members of that group do not share internal norms. This is the case, for instance, of The Dead. There are no internal norms that characterize the

dead since they cannot have them *qua* being dead. But there are widely shared external norms about them, like the norm expressed by the maxim "Do not speak ill of the dead" (more discussion in Kaspers, LiBrizzi, Calosi, Kobe, *under review*). But there also instances in which member of a social group do not identify themselves as members of a social group but such membership is imposed on them by external norms.

Consider this example. In Medieval times, society used to ostracize magic. In particular, this fear of magic manifested itself in the phenomenon of *witch hunt*. It might be the case that some people actually self-identified as *witches* (thereby sharing some internal norms), but the vast majority of people who was accused and executed for being a *witch* did not self-identify as a *witch*. Nonetheless, due to being accused of *being a witch*, they faced terrible consequences. *Witches* did constitute a social group, even though members did not share internal norms of self-identification or behaviour. Surely, we can say that the creation of the social group of *witches* was unfair and unjust towards people who fell in that group, but this does not mean that the group did not exist. If anything, granting the existence of such group helps us identifying such unjustness.

This is the reason why I endorse the claim that external norms are sufficient for the construction of social groups.

But let me briefly address a potentially controversial example. One way to read the debate about *transgender* people is that individual which are members of the group do self-identify as members of the group, but society has so far failed to recognize them as a group, thereby negating them the rights they deserve. Therefore, there are internal norms but no external norms. The question is: is *Transgender* a social group? Under my framework they are not, but they should be. We are not being unjust towards member of a social group; we are being unjust not to recognize them as members of a social group.

In conclusion of this section, I want to point out that, having argued that external norms are sufficient for the existence of social groups, we can extend the Normative Approach to social kinds as well. Of course, members of social kinds like *money* do not have internal norms. But we all have external norms about how to behave with respect to money. The difference is that such behaviours are not pointed at other people (like in the case of social groups), but at objects (broadly). Hence, The Normative Approach can account for social kinds as well as social groups, given the sufficiency of external norms.

5.4.5 Recap of the view

To conclude, let us recap the view and highlight some issues.

The Normative Approach, developed by Thomasson and discussed in the paper that I co-authored, says that social groups are characterized by norms. There are internal norms (guidelines for members of the group for how to behave with respect to themselves), structuring norms (guidelines for members of the group for how to behave with respect to other members) and external norms (guidelines for non-members for how to behave with respect to members of the group).

Under my interpretation, we have said that external norms can come from cultural heritage, moral principles, and/or new attitudes towards new circumstances. We have also seen that in order for an external norm to be widespread enough, it must be the case that either someone behaves in accordance with them (either consciously or unconsciously) or refuses them consciously (being aware of repercussions). Then, we have seen that people might have in mind different versions of the norm, but as long as everyone behaves in similar ways, that is not a problem.

Next, we have seen that, while internal norms are not sufficient for the existence of social groups, external norms are.

I think that there are two very positive aspect of this view. The first is that it is not vulnerable to the issues of *collectivity* and *acceptance* that we have seen in the Constitution Account. Recall that the problem of *collectivity* concerns the implausibility of all members of a community having active mental states about the constitutive rules. In this framework, the problem would translate in the implausibility of everyone being aware of external norms. This, as we have seen, is fixed by the possibility of behaving accordingly to the norm without actively knowing it. Norms are built into society, and it is to be expected that people respect them without being fully aware of them.

The problem of *acceptance* concerned the doubt about whether constitutive rules should be constantly accepted, or just *una tantum*. In this framework, the problem would concern whether external norms should be respected constantly or not. The answer is yes, they are respected constantly, yet this does not require active mental states and awareness.

The second positive aspect of the view is the adaptability with Reality Engineering. The normative approach provides a clear account of where defectiveness lies in our social world: defectiveness lies in the external norms. In everyday terms, defectiveness lies in how we behave with respect to one another. Discrimination, oppression, violation of rights, and other bad phenomena are grounded in the fact that people are used to behave in an unjust way towards member of certain social groups.

social groups exist, people do act in special ways towards other people *qua* being members of social groups, and these special ways of acting might be bad. The key aspect that the Normative Approach helps us to appreciate is that this normative

structure that surrounds social groups *seems* legitimate because it is widespread. This relates to the mistake of naturalization for social kinds. Social kinds seem like natural kinds because we are influenced by prejudice in saying that they pick out mind-independent facts about people and things.

But as we now know, social kinds are Constructed. This means that we can Re-Construct them, or Engineer them, if they turn out to have undesirable features. The Normative Approach tells us that if we want to modify the defectiveness of social groups, we have to modify our actions, behaviours, and practices. Most of my account of Reality Engineering relies on this: engineering is a matter of practice, not theory.

Chapter 6 – Extrapolating Two General Theories of social kinds

In the previous sections, I analysed four different theories of social kinds. As we specified many times, all of these theories belong to a general theory about social Phenomena known as social constructivism. Just to recall, social constructivism is the view that social kinds (but also social groups, Facts, Phenomena, etc.) do exist, and that their existence depends on our attitudes, practices, uses, cultures, linguistic utterances, and so on.

Each view has its flaws, as it is always the case in Philosophy. My point from the beginning of the thesis, however, is not to dwell too much on the single problems of each view, but rather to look at the practical implications of our theories. Such practical implications can help us understanding the social significance that social kinds have and how we can proceed when we try to modify them for the better.

My goal in this section is therefore to extrapolate the most practically significant features of the four theories previously analysed. In this process, I will highlight that the views that we analysed point us on two different directions for what concerns the nature and significance of social kinds. In particular, the Constitution Account and the Conferral Account seem to suggest that social kinds are a matter of stipulation, expertise/power and rules. While, on the other hand, the Entrenchment Account and the Normative Approach suggest that social kinds emerge from culture, common knowledge, widespread norms and behaviours and so on.

The key difference between these two patterns can be measured with respect to the general definition of social constructivism. We often defined social constructivism as the view that states that social Phenomena exist due to *our* practices (uses, attitudes and

so on). The two sub-patterns answer in different ways to the question of *whose* attitudes really count in the grounding of social Phenomena (kinds, in this case).

We can therefore define a Top-down pattern of views of social kinds as the group of theories that state that only a restricted group of people's attitudes and practices are relevant to ground social kinds, like in the Constitution and Conferral Accounts.

Similarly, we define a Bottom-up pattern of views such that most people's attitudes and practices are relevant to ground social kinds, like in the Entrenchment Account and the Normative Approach.

This division is extremely relevant because these two patterns differ substantially for what concerns the social significance of social kinds and what we can do to change them. But before we dive deeper into these two patterns to understand their general features, it is important to highlight that none of these patterns should be considered as exhaustive omni-comprehensive theories of social kinds. On the contrary, I wish to highlight that both patterns can and should be used to account for different social kinds which are currently at work in our society. I hope to show that some kinds are created and applied on society starting from the decisions and attitudes of a restricted group of individuals, may them be politicians, scientist, or else. Some other kinds, I believe, emerge from collective attitudes and widespread norms. This means that, depending on the kind in question, the metaphysics will be different and the way we should try to engineer it differs as well.

Let us then look at the two patterns.

6.1 Top-down social kinds

The Top-down pattern of views for social kinds is characterized by the idea that only a few people's attitudes and practices really matter in the construction of social kinds.

We have seen this idea at work in the Constitution Account (by Searle and Thomasson) and in the Conferral Account (by Ásta). My goal in this paragraph is to extrapolate the most important features of the two accounts in order to delineate the guiding principles of a Top-down view of social kinds. I will argue that there are three pillar concepts in Top-down views: Acceptance, Authority, and Application.

Acceptance, as we will see in a bit, concerns the stage where some individuals (I will explain who in the next paragraph) agree on some Constitutive Rules that apply a *social function* on an entity, which need not be a physical entity. By doing this, a prototype of a social kind is created.

Authority concerns the question of who in involved in the process of agreement over the Constitutive Rules. The concept of Authority is here to be intended as rather loose. Since in every community, even very small ones, different social kinds are at work, then in each of these communities different individuals will count as authorities. For instance, in social kinds that are at work in a State, it is likely that political figures have the role of authority; in an academic community, experts count as authorities; in families is likely that the elders have this role (more on this later).

Application concerns the stage where the proto-kind that the authorities agreed upon becomes a legitimate social kind which is operative in the relevant community. This stage is extremely hard to analyse since it is rather unclear how mass-beliefs and attitudes emerge. Perhaps, this would be work for sociologists, it is of no surprise that philosophical analysis fails to account for these matters in a convincing way.

Now, I will talk more in depth about these three pillar concepts. Afterwards, I will address the idea that this pattern of views is characterized by two different, yet equally relevant, stages: a Genealogical Stage and an Operative Stage. The Genealogical Stage

concerns the creation of the prototype and its application, which makes it a social kind. The Operative stage begins after the social kind is successfully established in the community. At this point, I will argue, what was in the beginning a subjective entity profoundly linked with the attitudes of those who agreed on its constitution, becomes an objective entity. In the discussion of this stage, we will get to rediscuss the idea that social constructivism does in fact account for perfectly objective entities.

- **"A" for Acceptance**

In accordance with the Constitution Account, the Top-down View states that the primary source for social kinds is the acceptance of constitutive rules. There are two main questions about this point: the first concerning the nature of constitutive rules (are they semantic rules? Are they metaphysical rules?), and the second concerning the nature of acceptance (is it an active mental state or can there be implicit acceptance?).
About the first question, one of the flaws of the Constitution Account, in my opinion, is that it does not really make clear whether such constitutive rules are semantic rules or metaphysical rules. I believe that we should have both semantic and metaphysical rules, but it is useful to distinguish them, especially if we are concerned with engineering projects (Semantic Engineering and Reality Engineering are substantially different enterprises). There is no doubt that the primary function of these rules is a metaphysical one. For instance, in the Constitution Account, the rule "x counts as y in context C" does not merely express a linguistic fact, but rather a metaphysical one: the grounding entity x, when the conditions of context C are fulfilled, gains a social function that makes it a different (social) entity y. Similarly, in the Conferral Account, what is conferred is a *property*, and the grounding entity becomes a new and different (social) entity. In both cases, constitutive rules work at the level of entities, not just language.

However, I do believe that we need semantic rules as well. Given that we are creating a new social entity, we must set in place rules that tell us how we are to refer to this new entity. The interesting point is that there might be feedback loops between metaphysical rules and semantic rules. By this, I mean that semantic rules are influenced by metaphysical rules, since semantic rules should try to delineate ways to refer to an object coherently with the nature of the object itself. But semantic rules say more than "x counts as y in context C". They might add specifications that are necessary for reference. These specifications might impact how the object is conceived and referred to. And since we are still at a stage where the metaphysics of this new social entity is malleable, the metaphysical rules may evolve following these ways of reference.

About the nature of acceptance, in general Top-down Views require explicit acceptance. We have seen this in the Conferral Account. In such case, the group of subjects who confer the property must have an active mental state about conferring that specific property to a specific entity.

We have also seen that the nature of acceptance was a relevant problem for the Constitution Account since neither Searle nor Thomasson really addressed the point. One option for them is to concede that acceptance can be implicit, more like non-resistance: if you do not actively reject something, then you accept it. In this case, they can reject Ásta's idea of Authority, and they could argue that all subjects within a community must accept the constitutive rules in order for them to work. The other option is to maintain that acceptance is an active, explicit, mental state. In such case, it would be too unplausible to claim that all subjects within a community have this explicit mental state. My speculative hermeneutics of this matter is that Searle would probably side with the first option, while Thomasson would go for the second. I say this

because we have seen that Thomasson's position has changed in time and in her later paper she argues for the Normative Approach, which can be seen as a response to the problem of Acceptance (more on this later).

- **"A" for Authority**

Top-down Views generally maintain that only a few peoples' attitudes and beliefs are relevant for the creation of social kinds. As we have seen, the biggest problem of the Constitution Account, in my opinion, is that it relies on an overextended notion of acceptance. Even though Searle and Thomasson did not really concern themselves with this issue, I believe that it is relevant. It is implausible that all subjects in a certain community have the same role in the constitution of the social kinds that count in that community and that everyone's acceptance is required.

Therefore, I think that the view is better off maintaining a non-collective notion of Acceptance. The choice of the word Authority recalls the Conferred properties framework. However, my concept of Authority is different from Sveinsdóttir's.

Again, I believe that the Conferred properties framework is guilty of some oversimplification. Indeed, in the examples provided, Sveinsdóttir takes one subject as the authority that confers a certain social property, like the referee in baseball. Also this view seems misleading.

This is why I propose a concept of Authority which is not as restrictive as in the Conferred properties framework, and yet not so permissive as in the Constitution Account. In the Top-down pattern of views, Authorities are going to be a group of people with specific competences and mansions. I believe that this concept comes very close to the typical concept of Expert, which is diffused in discussions in Epistemology. Clearly, some work would have to be done in order to define who counts as an expert,

since there is the risk of falling into circularity. I will not analyse this problem further, since I am not trying to provide a fixed view but a pattern of view; the point is that each view in this framework has to account for this issue. But one more thing can be said at this stage.

In the case of social kinds, the figures of experts are more complicated than in the case of natural kinds. For instance, Physicists (people with a very specific preparation) are in charge of defining and shaping Physical kinds (same for Biologists, Zoologists and so on). On the other hand, it is much harder to find who counts as expert in the case of social kinds. Maybe in some fields, like Economics, we can apply the same standards that we used with natural kinds (Economists, who have a very specific formation, should be considered experts). But what about kinds in the Gender or Race areas? Starting from this issue, I will start distinguishing the Genealogical side of the Top-down views from their Operative side.

In particular, if we look at the genealogy of kinds in the Race area, we see that people who introduced the various categories were "illegitimate authorities", in the sense that they did not have any competence in the area and they should not have introduced such categorizations.

This is extremely relevant for Engineering projects, since one of the prominent ways in which a Top-down social kind can be defective is that the authorities that stipulated it were "illegitimate authorities". Considering an ultra-contemporary instance, it is not up to politicians to define the usefulness or relevance of vaccinations. It would be up to doctors and scientists to do so. Therefore, it is of no surprise that many people reject vaccinations, they fear that this is a political matter when it is absolutely not. If the

communication of the nature and usefulness of vaccines was left to proper experts only, I believe that many more people would have agreed to it.

- **"A" for Application**

The application stage is the one in which the prototype of social kind which was formed via the Acceptance by the Authorities confronts itself with the world. The mechanisms of the Application stage are complicated, and somewhat mysterious. Sometimes, the process in which a kind becomes part of the operative social kinds in a certain community obtains or fails for opaque reasons. But sometimes it is clear why the kind fails to obtain: the category may fail to track a unified group of things/people in a certain community. For instance, a prototype of Human kind could unify a group of people with features and behaviours which are too different from each other. In this case, we may say, the social World rejects this proto-kind, and it never ends up being an operative kind in the community.

Also, addressing one point raised during the discussion on Acceptance, the semantic rules and the metaphysical rules could respond in different ways during the application stage: the group of things/people could end up being part of an operative kind which has a different name or different semantic rules. Vice versa (more likely), it could be the case that the semantic rules obtain and end up being operative while denoting a different group of things/people than the one defined by the original metaphysical rules. Here is where many cases of defectiveness of social kinds emerge.

One last consideration: in the application stage we lose Epistemic Transparency. Since the operative kind may turn out to be significantly different from the proto-kind due to the confrontation with the world, such operative kind can exist even if no one believes it to exist in that form.

- **The Genealogical Model**

The Top-down pattern, as it should be clear by now, is divided in two crucial stages. First, we have the Genealogical stage, which consists in the creation of a proto-social kind. The second one, the Operative stage, is the one that follows the process of application, in which the social kind is a fully-formed operative component of the social World.

As a quick overview, I will sketch out the Genealogical story of Top-down Views.

As we said in the previous paragraphs, the process of creation of the proto-social kinds happens via the acceptance of semantic and metaphysical constitutive rules by some authorities (or experts). The bit of the story that remained untold is the one concerning what grounds the expert's choice and what the constitutive rules are about. I also anticipated that I would take distance from the mere typical form "x counts as y in C", it is now time to explain why. It is quite clear what "y" and "C" represent in the previous formula: "y" stands for the proto-social kind that is being created, for instance "x counts as money in C"; "C" stands for the context in which such social kind is projected to operate, for instance "x counts as money in Italy". However, I find difficulties understanding what "x" is supposed to stand for. When we look at individual coins or bills, it seems natural to say that "this coin counts as money in Italy". But here we are committing the mistake of mixing the Genealogical and the Operative stage. Such sentence, in fact, is uttered when we are already in presence of an operative social kind and we state that a certain individual is part of it. In this case, in order to measure whether the individual coin is part of the kind we must start from a starting standard, or definition. But in the Genealogical stage we do not have the definition, we are creating it.

According to Top-down views, the process of creation of a proto-kind starts with the observation of similarity relations between various individual entities in the physical (or social) world. Let me clarify by taking an example from the natural kinds case. Suppose you are a physicist in 1897, when we did not know about Electrons. During your experiments, you start observing that some particles behave in a different way than the others and their behaviour is quite homogeneous. Also, you observe that they share other features, like the value of their mass. At this point, you give a common name to all these particles, and you call them "electrons". You just created a proto-kind. In this case, the constitutive rules would be something like the following: "all those particles that share features f1, f2, …, fn and share behaviours b1, b2, …, bn are *electrons*". Similarly, consider the following case. You observe that certain people have similar behaviours, like being particularly addicted to technology and avoid social gatherings. Also, you observe that they also share other less notable behaviours, like dressing similarly. You give a common name to these people, and you call them "nerds" (let us ignore the initial negative connotation that these terms actually had).

The most important aspect of this view is that the starting point for the creation of a proto-kind is the observation of similarity relations between individual things or persons. This is why Top-down Views, at least for what concerns the Genealogical stage, is an Extension-First View. By Extension-First I mean that in the creation of the social kind we start from observation of individual entities, and not from a pre-determined definition. Oppositely, in Intension-First views, you assign membership to a kind by measuring whether the individual thing/person meets the standards of the definition of such kind. This is unfeasible as genealogical explanation.

At this stage, it is interesting to look at the role of subjects and objects in this genealogical process. By baptizing a social kind, experts are creating it as a proto-entity of the social World. This means that experts have a metaphysical role in the genealogical stage of Top-down views. At this point, other subjects of the community do not have any role since they are not part of the creation of the proto-kind and such proto-kind did not enter the scene of the social World.

Oppositely, at this stage, entities in the world have an epistemic role. As in the Conferred properties framework, the world is just the content of some judgment by the creators of the proto-kind, and it does not play a role in influencing the structure of such proto-kind.

- **The Operative Model**

The Operative stage of Top-down views starts when a proto-kind successfully surpasses the process of application. At this stage, the emerging social kind is a fully formed entity that legitimately enters the landscape of the social World. At this point, most of subjects in the community are aware of the existence of this kind (even though they might have wrong beliefs about it). An interesting aspect of the Operative Stage concerns the evaluation of potential new members of the kind (or of members that lose their status).

Contrary than in the Genealogical Stage, during the Operative stage we start from the definition of a certain kind in terms of which properties something must have in order to be considered as member of the kind, and we analyse the potential new member in order to see whether it does indeed possess those properties. The Operative stage, therefore, is an Intension-first stage. Similarity relations become less important and all that matters is the respect of the definition (i.e. the Intension). This, I think, is particularly interesting,

because it inverts the roles of subjects and the world with respect to those that they had in the Genealogical stage.

Subjects do not have a metaphysical role over the kind any longer, the kind has become an independent entity by passing the Application stage. A subject's judgment is not enough to modify the kind (lots of discussion to come on this point in the future of this research). Subjects become users (or consumers) of the kind, no one (not even authorities) is a creator any longer.

Oppositely, the world, the social World in this case, has a metaphysical role. The metaphysical structure of a kind determines which members may be added or subtracted from it. Now the kind plays an active role on our beliefs, practices, and judgments: the kind now determines, in part, the content of such judgments.

One last quick point: the distinction between the Genealogical and the Operative stage allows us to avoid the process of naturalization of social kinds: if we are able to tell the story of the creation of a kind then we see that its metaphysical structure is not determined by intrinsic features of the world but by us. The strong ontological status of the kind that survived Application should not be considered as a sign that the kind accurately reflects the intrinsic structure of the World.

- **Room for plasticity**

In conclusion, I would like to highlight a couple of potential spots for plasticity within Top-down Views.

First, social kinds in Top-down views are intentionally-free plastic: they can change regardless of human actions directly intended to making such change happen. It is a fact that the social World is in constant evolution. The constant change of the social World in general may lead to the modification of individual kinds as well. For instance, a

certain community could be influenced by a foreign culture with other values and standards, and our social categorizations may unconsciously evolve towards these new influences. New authorities might come up for independent reasons and they may stipulate new kinds due to these influences. Or the application stage could be favoured or slowed down due to some external reasons and influences. Or there might be events that change our priorities and categories. From a theoretical standpoint, however, intention-free plasticity is to be explained in the following manner. If the process of application were a one-time phenomenon, after it is completed, the kind would not be influenced by anything external to it. This is why we have to extend the process of Application: after the first phase, which we might call Establishment (of a certain kind as a legitimate entity, through confrontation with the world), there is a phase of Continuous Application, that allows the kind to respond to other entities and events of the social World.

Lastly, and most importantly, intention-bound plasticity. In Top-down Views, social kinds are intentionally-bound plastic: they can change because of human actions directly intended to making such change happen. In particular, we might modify who counts as an expert and whose acceptance counts in the process of creation of proto-kinds. Also, we might find defectiveness for what concerns the application of a certain proto-kind: the result might be too distant from the original prototype. Or the kind might be similar to the prototype but defective from a moral, social, or political standpoint. In all these cases, and in many others (which I will analyse later on in the section about Engineering), we might be able to open the door for Reality Engineering.

6.2 Bottom-up social kinds

The case of Bottom-up social kinds is rather different than the one of Top-down kinds. Here the key is no longer the acceptance (as an active mental state) of some constitutive rules. The idea is that Bottom-up social kinds emerge from practices, attitudes, uses, linguistic utterances, habits and beliefs that are widespread in the community where the social kind is at work. In short, we might call these practices social norms. As a consequence of this definition, we may say that Bottom-up social kinds are nothing but the reification of such social norms.

Many questions arise from this simple definition, questions that I will attempt to answer in this section. First, we might ask what are social norms, what is their form, and what it means for an individual to behave in accordance with these norms. Then, we might want to know what is a reification of these social norms, thereby explaining what social kinds are. Also, we might as how widely shared these norms should be in order for them to constitute robust social kinds. Furthermore, there are other questions concerning the grounding of these norms (where do they come from? Are they legitimate?), what it means to reject these norms, what happens if they are unfair to someone and how to change them or create new ones.

- **What are social norms and what grounds them**

The Bottom-up pattern of views of social kinds is a group that includes, together with other views, the Entrenchment Account, and the Normative Approach. In both of these views that we analysed above, the key element for the constitution of social kinds is to be found in people's attitudes, practices, beliefs and behaviours.

Mallon and Tsou call these *social roles,* while Thomasson calls them social norms. Let me briefly recap these two definitions.

According to Mallon and Tsou, social Roles are individuated by representations like "attitudes, theories, narratives, concepts, models, pictures, norms, rules, utterances" (Mallon, 2016, p. 6) and categories are defined as properties (intensionally) or classes of individuals (extensionally).

He points out two necessary existence conditions for social roles. A social role exists iff:

3. "There are representations that pick out a category of persons[5] and a set of beliefs and evaluations associated with that term" (Tsou, 2020, p. 117).
4. "Many of the beliefs and evaluations associated with the role are common knowledge within society" (Tsou, 2020, pp. 117-118).

Two elements emerge as fundamental from this definition: representations (and beliefs) and common knowledge. Surely enough, such representations and beliefs are connected with practices and behaviours, if we set aside the rare cases in which a person acts in opposition to her own beliefs.

On the other hand, Thomasson's idea of norms has a more directly behaviouristic connotation. She divides between internal, structuring, and external norms:

internal norms are "norms regarding how members of the group are to behave, regard themselves and other group members" (Thomasson, 2019, p. 4838). These norms give guidelines to people who are *inside* the social groups about which attitudes and behaviours are appropriate to have *qua* being members of the group. Such attitudes and behaviours are specific to single groups, and the only reason that one must have in order to follow them is to self-identify as a member of the group. To give an example, *as a member* of a basketball team, I stand by the norm of showing up to practice every time.

For me, there is no other reason for showing up at practice if not for my membership to the team.

structuring norms are "norms that place different members of the group at different nodes, with different norms regarding those who occupy different nodes" (ibid.). In cheap terms, structuring norms tells members of groups whether there is a hierarchy of some sort built into the group, and, if so, what attitudes and behaviours a member of one of the layers of this hierarchy should have with respect to members of other layers. As instance, as a student who is *member* of the social group that is the University, I am aware of how I am expected to behave with respect to my supervisors and other faculty members.

external norms are "norms regarding how members of that group are to be treated, regarded, behaved towards by those who are not members of the group (ibid., p. 4839)". An easy example is the one of the social group of the *elderly*. We all (mostly) stand by the principle that we should treat the elderly with respect. This is exactly what Thomasson means when talking about external norms: they are sets of behaviours and attitudes that people are expected to have towards members of social groups.

As can be noted, there are not too many differences between the two views, except for linguistic ones. The idea is that, fundamentally, social groups are the expression of sets of beliefs and behaviours that are widespread in society.

As I have argued above, there are good reasons to believe that external norms, as defined by Thomasson, have explanatory and ontological priority over internal norms: internal norms are not necessary nor sufficient for the constitution of social groups, external norms are.

This reasoning allows us to expand the scope of these two definitions to social kinds as well as social groups. Members of kinds that do not have humans as individuals do not share internal norms[6]. But since internal norms are neither necessary nor sufficient for a social kind to exist and external norms are, kinds that have members which are incapable of sharing internal norms should be considered as social kinds nonetheless, as long as external norms are in place.

One other key point about social norms is the question of their origin. Where do they come from? What legitimates them? I will argue that there are three prominent ways in which these norms emerge: tradition, response to new circumstances, engineering old norms.

We have countless examples of social norms that are at work in contemporary societies which come from culture and tradition. For instance, norms about treating the elders with respect, about behaving appropriately in the presence of certain people, about the distribution of political responsibilities in certain groups, and many others. There are also, and this is a fundamental point, norms that come from tradition which are not just nor justifiable on other grounds (moral or social). These are norms that lead people to discriminate others on the bases of race, sex, sexual orientation, religion. Another instance is represented by the ways in which we treat animals. According to most cultures and traditions (except, perhaps, for some cultures in the area of India and Nepal), it is completely accepted that we exploit, torture and kill animals. Surely, at the time when these behaviours emerged, people needed to eat animals to survive. But nowadays we know for a fact that we do not need to eat animals to survive. Therefore, the only moral and social justifications for the horrors that we commit towards animals

is solely the one of culture and tradition (other than taste). It should be evident that sometimes such justifications are not good at all.

Another way in which social norms emerge is due to a community's response to a new set of circumstances. New circumstances which are not covered by social norms of the past lead to the creation and diffusion of new attitudes, beliefs, and behaviours. We can call this a form of *de novo* engineering, and I will address this further in the section about Reality Engineering projects with respect to Bottom-up kinds.

The third way is *de vetero* engineering: the modification of old social norms that are no longer considered just or justifiable. I will discuss this further in the section about Reality Engineering as well.

The last point about the nature of social norms is to address the difference between them and the social Rules typical of Top-down social kinds. Let me highlight that this difference is due to my own definitions, and might not reflect everyday usage of these terms, this is a framework-relative discussion. So, in this framework, there are two points of divergence.

The first is that social Rules apply a *social function* to some entities, so that the relevant social property is in the object, not the observer. Social norms, on the contrary, are such that the relevant social role and property are in the observer, not in the object. This will be clarified in the next section, but as an anticipation we can say that Bottom-up social kinds are not grounded of properties of their members, but rather on properties of external individuals, in terms of their beliefs and behaviours towards the internal members. Top-down kinds, on the other hand, are such that social properties are possessed by the internal members.

The other relevant point of divergence between Rules and norms in this framework is that Rules are shared between a small portion of the relevant community while norms are shared between a large portion of the community (more about this in a bit). Surely enough, in the Operative stage of Top-down views, members of the community share social norms, since they all have beliefs and attitudes towards the social kind in question, but this does not conflict with my argument, since the grounding elements of Top-down remain Rules, and not norms.

- **The metaphysics of Bottom-up social kinds**

Bottom-up views of social kinds have their key feature in the notion of social norms. According to these views, social norms ground social kinds, not only from an explanatory perspective but also from a metaphysical perspective.

social groups are grounded primarily on external social norms, since the beliefs and attitudes of individuals who are external to the group shape the intensional definition of the group, and (as a consequence) its extension in terms of members. Internal norms still play an important role, since the attitudes and beliefs of a certain group may influence people outside the group to behave towards them in certain ways. It may happen also that internal norms take place temporarily prior to external norms. For instance, the social group of *vegans* is such that internal norms were born before external ones: a group of people started showing the pattern of behaviour of limiting to the maximum the consumption of animal products, and they identified themselves as a group. But only when people started recognizing them, they became an actual social group.

For a social group to exist it must play a role in society, and to play a role in society it must be recognized by other members of the relevant community. However, if we were

to delineate the features of the social group of *vegans* in terms of their role in society, we would rather say that they are seen as extremist than as people who make a noble moral choice. Internal norms make members of this group believe that they are indeed bearers of good moral values, but external norms make non-members of this group believe that the internal members are extremist, pushy and unreasonable. The point is that from an economic, political, and moral perspective, the norms that count are the external ones, so that little to no effort is made to facilitate a vegan lifestyle.

So, ultimately, it is the external norms that ground the actual social group of *vegans*.

social kinds are less controversial, since their members are not individuals who share internal norms, but entities which do not have beliefs and intentions. Bottom-up views have an easy game in arguing that external norms ground them, since there are no other norms in place.

Some Bottom-up views could push this idea of grounding even further, as far as arguing that social kinds do not amount, from a metaphysical perspective to nothing else than the norms that dictate behaviours towards their members. By metaphysical perspective, I mean that the only properties which are relevant for the definition, constitution, and existence of these kinds are those emerging from the relevant norms. In other words, these views state that the intension of social kinds comprehends only those properties that come from the social norms affiliated to the relevant kind.

I do not aim to defend this point here, but I think we can take it as the starting point for an useful discussion. At this point, we might ask whether Bottom-up views are Intension-first or Extension-first. That is, do definitions in terms of properties have ontological priority or do definitions in terms of individual members have such priority?

I think that Bottom-up views should commit to an Intension-first perspective. This follows from our discussion above. The case of social kinds is again uncontroversial: widespread social external norms shape and define the social kind in question, and this definition is in terms of the social properties that the members of the kind are believed to possess (and that, therefore, they actually possess). The non-social features of members of the kind do not play a relevant role in such definition. This means that the Extension (in terms of the members) is ontologically secondary with respect to the Intension.

The case of social groups again depends on the ontological privilege that external norms have over internal norms. If internal norms are secondary, it means that internal features of members of the relevant social group are secondary to the social properties that are projected on them. The latter properties are those that constitute the intension of the social group. Therefore, the Intension is once again ontologically prior to the Extension.

- **How widespread norms are and the role of dissidence**

One last fundamental consideration about social norms concerns their diffusion: how widespread do these norms have to be in order to constitute legitimate and operative social kinds? In other words, what attitudes should people in a community maintain with respect to these norms in order for them to constitute social kinds?

One might think that since Bottom-up views maintain that social kinds emerge from the social texture due to the existence of patterns of beliefs and behaviours known as social norms, these social norms should be respected and known by every member of the community.

This view, however, is untenable. It is factually false that all members of a community where a social kind is at work have beliefs and attitudes towards members of such kind

(may these members be people or entities). This means that not all members of a community actively and consciously respect the social norms affiliated to the kind.

On the other hand, Bottom-up views cannot hold that only a few people in the community share social norms, because that would make the view collapse in the Top-down pattern. Therefore, we need something in between.

The most neutral way to express the necessary diffusion of social norms is to say that for every individual in the community either she accepts the norms actively and behaves accordingly, or she behaves accordingly to it without knowing it, or she rejects it consciously. Only one option is not viable in this framework: you cannot unconsciously reject a social norm without facing consequences. This idea comes from Thomasson's insight, when she says that:

As participants in the social world, we can be blind to these norms and fail to recognize the social groups […] only at our own peril — peril of sanctions, ostracism, or worse. (Thomasson 2019, 4840)

Therefore, the threshold that a social norm has to surpass in order for it to be considered widespread enough is that whoever rejects it is passible of negative consequences. Whoever rejects the norms, must know the possibility of such consequences.

This does not mean that social norms should be (and are) respected by everyone. Dissidence is to be expected, and it is a desirable factor in our social world. We cannot have a society where everyone behaves in the same way and thinks in the same way. What we need to understand is the role of dissidence in this framework, which is to understand what it means to reject social norms and what impact this dissidence has on the norms themselves. I will divide the analysis of this matter in two parts since (at least in the case of social kinds that have human individuals as their internal members) there

are two groups of people whose attitudes are sensibly different: internal members and external members. The ways in which internal and external members reject social norms are relevantly different and they have different effects in our framework, so I will discuss them separately.

Rejection of social norms and dissidence by internal members of social groups (kinds that have humans as members) can happen in two ways. First, one might not identify as a member of the social group that external norms project on her. Second, one might identify in the group but reject the features that are affiliated with the group.

As an instance of the first you might think of a young person who was born with the biological features typical of the feminine sex and is considered as a girl by everyone. They, however, do not identify as girl or women but as a boy. In this case, this person refuses the membership to the social group that is imposed on them *tout court*.

As an instance of the second you can think at women who identify as women but refuse the etiquette that society imposes on them, that refuses to be considered somehow less worthy and discriminated. In this case, these people do not refuse norms in terms of their identification as women, but they reject and criticize the features that social norms impose on them, i.e. how they are treated.

The problem with internal dissidence, however, is that even if the totality of members of a social group were to reject the norms affiliated with the group, this by itself would not lead to a modification of these norms. This follows from the previous discussion: internal norms are not sufficient for the existence and the ways of existing of a social kind, external norms are. Furthermore, it is often the case that internal dissidence leads to bad consequences for those who reject the norms. Whether by further

discriminations, reinforcement of norms or other ways of repression of dissidence, it depends on the case and context.

And if this sounds weird and extreme, think of an ultra-contemporary example. Since the comeback of the Taliban regime in Afghanistan, we have heard that women are going to lose most of their rights, like the possibility of participating in political matters and even the right to education. Basically every woman is against the new norms that are likely to be imposed on them (except, maybe, for a very small minority). They all reject the external social norms that men have towards them, they all reject the status and features of the social group of *women* which is at work in their society. But this, by itself, is not going to stop these new norms to emerge, and it is not going to change the operative and actual status of women there. The prevalence of external norms is crude and often problematic, but this does not make it less true.

This factor is going to be key in our development of Reality Engineering: dissidence usually arises internally to social groups but is only when internal members can communicate and convince external members that they are being oppressed or unfairly treated that things can really change.

Dissidence of individuals who are external to a social kind or group is very different. First, by definition, rejection and dissidence towards external norms has to be minoritarian. If it was the case that most members in a community refused certain external norms, such norms would cease to exist, since external norms are widespread by definition.

However, dissidence has to be expected in a democratic society. Such dissidence can be manifested by merely ignoring the norm, by actively rejecting it and by campaigning for its rejection. As long as dissidence remain a minoritarian phenomenon, those who reject

the norms are subject to bad consequences, like, as Thomasson says, “peril of sanctions, ostracism, or worse” (Thomasson 2019, 4840). However, if external dissidence were to become sensibly widespread, that would lead to the process of modification of the social kind or group affiliated with the norms that are being contested. This is going to be the subject of discussion of the section about Reality Engineering of Bottom-up kinds, and I delay this discussion until then.

- **Room for plasticity**

As in the case of Top-down social kinds, I want to briefly discuss the potential role of plasticity in the framework of Bottom-up kinds. The question to analyse is whether the theory can account for social kinds that change over time while maintaining their identity, and whether such change is the result of human actions directly intended at making it happen (intention-bound plasticity) or it is the result of some other factor (intention-free plasticity).

First off, Bottom-up social kinds are intentionally-free plastic. This is motivated by the fact that social norms change all the time in society, and such change is not always motivated by actions directly intended at making it happen. For instance, some environmental factor or the confluence of different cultures can help people behave differently towards certain social groups or kinds. It might very well be the case that neither the person whose behaviour changes nor anyone else has made actions to provoke some behavioural change (i.e. to make the social norms change).

Bottom-up kinds are also intentionally-bound plastic. In this case, arguing for intention-bound plasticity is simple, thanks to the practice-first approach of this account. Since social norms are a matter of beliefs, practices, and behaviours and all these things can be changed and acted upon, it is clear that intention-bound plasticity holds. In front of

some kind of defectiveness in the social norms at work, we might regulate our behaviour and engage others to do the same. A discussion of how this can be done is deferred to the section about Reality Engineering.

Chapter 7 – Reality Engineering for Top-down social kinds

In the previous chapter, we have taken a look at two general patterns of views for social kinds, one that conceives social kinds as the result of the application of a stipulation by some authorities in terms of *social rules*, which we called Top-down pattern. We have seen instances of this pattern in Searle's and Thomasson's Constitution Account and in Ásta's Conferral Account.

The second, which we called Bottom-up pattern, conceives social kinds as emerging from *social norms*, specified in terms of beliefs, practices, behaviours, uses, linguistic utterances and so on. These social norms have to be sufficiently widespread in a community for a social kind to be at work in that community. We have seen instances of this pattern in Mallon and Tsou's Entrenchment Account and in Thomasson's Normative Approach.

We did not advocate for one pattern over the other as the best explanation of social kinds. Even further, I have suggested that whether some social kinds at work in our society are best explained by Top-down views, others are best explained by Bottom-up views. The two patterns can be complementary and exhaustive, depending on the connotation that views within the pattern can take.

In this chapter, I will focus on an analysis of projects of Reality Engineering for Top-down social kinds. First, let us recall the key features of Top-down social kinds, since this is going to be key in the understanding of what we have to do in our enterprise to modify their defective features.

We have said that there are three key concepts to bear in mind while discussing the nature of Top-down social kinds: Acceptance, Authority and Application.

- Acceptance concerns the key moment of creation of a proto–kind. This happens when a restricted group of individuals within a community stipulate and accept some *constitutive rules* that impose a *social function* on some entity, which need not be a physical entity. These constitutive rules can be both metaphysical rules and semantic ones, where metaphysical rules say that the grounding entity x, when the conditions of context C are fulfilled, gains a social function that makes it a different (social) entity y. Semantic rules, on the other hand, specify the conditions for reference to the social entity y. If the process of Acceptance is properly brought on, we will have both metaphysical and semantic rules. Furthermore, acceptance in this framework is to be considered as an active mental state retained by those involved in the process of stipulation and agreement on the rules.
- Authority concerns the question of who are the individuals whose acceptance matters in the creation of the proto-kind. In the instances of the Top-down pattern of views (Searle and Thomasson's view and Ásta's) we saw some not convincing answers to this matter. Searle and Thomasson's view was too liberal, such that too many individuals were supposed to participate in the process (which collides with the idea that acceptance should be an active mental state retained by everyone involved). Ásta's view was too restrictive, such that even one single subject could fulfil the role of authority, like a referee in baseball. Therefore, I suggested a mid-way view where authorities are conceived similarly to Experts. As we will discuss in a bit, who counts as expert is controversial, and one of the main causes of the defectiveness of Top-down social kinds.

- Application concerns the stage where the proto-kind stipulated by the Authorities via the acceptance of constitutive rules is projected on society. We discussed that it is often unclear how such stage actually works, but we can indeed identify some ways in which the process can go wrong, which is the feature that most interests us in Reality Engineering projects.

With this in mind, let us also recall what we said many times about the primary condition for a Reality Engineering project to be possible: in order to be engineerable, an entity has to be plastic, i.e. it must be able to change over time while retaining its identity. We have identified some room for plasticity in the Top-down pattern of views. We have observed that Top-down kinds can change because of human actions directly intended to making such change happen. In particular, we might act in order to change who counts as an expert (i.e. whose acceptance counts in the process of creation of proto-kinds). Also, we might fix problems in the application of a certain proto-kind: the result might be too distant from the original prototype. Or the kind might be similar to the prototype but defective from a moral, social, or political standpoint.

As a further preliminary note, we can easily notice how these potential defective features of Top-down social kinds are not to be found in their theoretical aspects, but rather on their practical developments. This observation gives us once more the chance to highlight that Reality Engineering is an extremely practice-oriented enterprise. What we act upon are pragmatic aspects of our everyday life and of our societies. I am open to the idea that the actual engineering side of Reality Engineering does not fall in the domain of philosophy. Philosophical analysis is here the tool that we use to analyse the structure of the entity that we need to change. By doing that we get guidelines for how

to direct our actions and what to focus on, but the actual work is purely a matter of practice.

I consider this a virtue of Reality Engineering, since, as we have seen, one of the main problems of Conceptual Engineering is that it presents itself as a purely theoretical enterprise and struggles to find practical applications.

Given that Top-down social kinds rely on a concept of authority, it is inevitable that the main applications of Reality Engineering in this framework are political ones, in contrast with the case of Bottom-up kinds that, as we will see, rely on a more behaviouristic and communitarian approach.

In order to get started with the analysis of how a project of Reality Engineering can be fulfilled, it is important to recall once again that what we are trying to do is to fix the defective features of a certain kind and replacing them with new, better functioning ones, while maintain the good features that the kind had in the first place in order not to lose its original function.

Therefore, depending on where we find the defective features of a certain kind, we have to perform different actions to fix them.

In order to clarify this point, I will start by analysing three possible cases of why a social kind can be defective, and I will give some suggestions for how to fix these different forms of defectiveness.

- **Case 1**: *The constitutive rules are defective*

There are different reasons why constitutive rules can be defective: the authorities might have agreed upon them for disingenuous reasons, or for their own interest. They also might have been not proper experts in the field, thereby causing poor rules due to their

scarce competence. Or authorities might have simply taken a mistaken, yet not malicious, decision.

Let us think about some examples.

Consider the social economical kind *tax* in a hypothetical state where taxes are distributed in a way such that poorer families pay 25% of their income in taxes and richer families pay 15% of their income in taxes. Such system is obviously flawed since it puts poorer families in a condition of unfair disadvantage and oppression. It is uncontroversial that the social kind of *tax* is defective in such state, and this defectiveness lies in the constitutive rules, since in this case the rules are those that state the distribution of taxation. Note that all three of the reasons for defectiveness could be at work in this scenario. It could have been the case that the authorities in question took this decision because they are part of those richer undertaxed families, thereby stipulating the constitutive rules for their own convenience. But it might also have been the case that such authorities had no real conception of how such system would harm poorer families (even if this is hard to conceive), thereby showing their lack of competence. Or they could have made a genuine mistake, by thinking that reducing taxation for richer families would lead to an increase in investments and profits, without considering the status of poorer families.

As another example, think about the social kind *grade* at work in a hypothetical high school where students' works are evaluated on a scale from 1 to 10 and a student can only take a grade which differs by 1 (plus or minus) from her previous grade. The system is obviously flawed, because students that have high initial grades are disincentivized to work hard, since in the worst-case scenario they will get just one fewer grade than the previous time. But this evaluation system is also unfair for students

that got lower initial grades, since they are never given the chance to improve their situation, even if they do a particularly good job. We can therefore say that the kind *grade* is defective in such high school, and this defectiveness lies once again in the constitutive rules at work in this scenario, since the rules and criteria for evaluation are the very reason of the unfairness of the system. In this instance, it is unlikely that the authorities (the committee board of the high school) acted due to their own interest since it is not clear what advantage they would get from such an evaluation system. They could, however, have made a mistake by thinking that such system would protect better students, avoiding that a single bad exam would impact their general evaluation too heavily. Such a benefit would not make up for the bad effects that we discussed.

What can be done in situations like these to fix the defectiveness of the social kinds in question? I believe that there are two main solutions.

- **Solution 1**: *Change the authorities*

Once philosophical and empirical analysis of the phenomenon in question show the nature and reasons of the defectiveness of a certain social kind, and this defectiveness turns out to rely on the very constitutive rules stipulated and accepted by the authorities, the first solution is to change the authorities who took the decision.

This step is necessary when we realize that authorities acted due to their own interest, or in other disingenuous ways, or because of their incompetence.

Changing the authorities at work in certain frameworks is an extremely difficult enterprise, as we will se at the end of this chapter when we discuss issues of feasibility for Reality Engineering projects. For instance, in the case of *tax*, for a change in the authorities we would need a change in the politicians who are in charge of taking such decisions. This process would obviously take time and effort.

The important point is, however, that philosophical and empirical analysis have the fundamental role of unveiling the defectiveness of our social kinds. Once that is done, we can convince others that such social entities are flawed and unjust and try to take measures against them.

Suppose, however, that we did succeed in voting a new government in the hypothetical state of the example of *tax*. At this point, we get new authorities. What happens then is a new Genealogical Stage: authorities have to rediscuss constitutive rules, stipulate new ones, and agree on the creation of a new proto-kind. Going back to Conceptual Engineering talk, this process resembles what *replacement* engineering does: it starts back from scratch and try to create a new and better concept, rather than modifying the old one[7].

- **Solution 2**: *Keep same authorities, change their mind*

The second possible solution to fix social kinds that are defective in their constitutive rules is to keep the same authorities and change their mind, thereby pushing them to modify the unfair kind that they created.

This solution is only possible when the authorities made a sincere mistake, without any malice or interest (because we would fall in the scenario of Solution 1).

Philosophical and empirical analysis can help identifying why a social kind, like *tax* or *grade* in the examples, is flawed. Such analysis should count as evidence for the authorities to realize the mistake and try to fix it.

For instance, the board of professors at our hypothetical high school, after having been made aware by empirical analysis of the difficulties that their evaluation system causes to worse students and of the disincentives that better students face and by philosophical analysis that the cause of these difficulties is to be found in the constitutive rules that

they themselves stipulated and agreed upon, should consider a re-evaluation of their constitutive rules for the kind *grade*.

They should look for new and better constitutive rules that do not discriminate between students that started their career well and those who had initial difficulties. Also in this case, we enter a new Genealogical Stage where the authorities look to create a new and better proto-kind to be then applied in their school.

- **Case 2**: *Application went wrong*

Another way for a Top-down kind to be defective is for the stage of Application to have gone wrong in some way. As we pointed out, the process of application is unclear and there are plenty of ways in which it can go wrong. The point is, however, that empirical and philosophical analysis can help us to point out at the factors that contributed to the failing of the process, even if we are not fully aware of all the features of such process.

To clarify, think about this example.

Consider Italy's strategies for vaccination against Covid-19. The social kind in question is the kind *vaccine*. This is a beautiful example of a Top-down social kind because we have a physical entity such as the vaccine itself upon which some authorities (the politicians) imposed a *social function* such that whoever is vaccinated is free to go to restaurants, museums, cultural events and so on. In terms of the constitutive rules, in the context of Italy right now, the physical entity 'vaccine' counts as a 'green pass' for people to go (almost) wherever they want.

In Italy, however, the process of vaccinating the population did not go as well as desired by the authorities who proposed. Millions of people decided not to get vaccinated, and some even actively protested against vaccination and its function as a green pass.

We might therefore say that the kind *vaccine* is defective in Italy, because not enough people are vaccinated, and the health of the population is not guaranteed enough.

It is debatable whether the very idea of vaccines counting as green passes for people to go anywhere is flawed, which would make the kind *vaccine* defective in its constitutive rules. It might be argued that the measure is too discriminatory, and that non-vaccinated people should be let free to do more things than they currently can. But let us suppose that there is nothing bad in the constitutive rules, and that it is indeed a good idea to restrict access to certain activity to vaccinated people.

Then, the defectiveness is to be found in the application stage of the kind *vaccine*. The good proto-kind created by the authorities failed to be a good operative kind in society, and really turned out to be a source of social conflict and discrimination. What went wrong there?

The main cause of this failed application is to be found in the bad way in which the authorities communicated their intentions and their motivations to the general public.

First, the communication was monopolised by politicians, not giving doctors and scientists enough institutional relevance. They made it look like vaccination is a political act, rather than an act directed solely at the preservation of people's health. Doctors and scientists worked on the background, displayed only in occasional tv programs.

Second, the authorities dealt with the problem of people who did not want to get the vaccination simply by blaming them, without giving real help to people who refused due to fears, prejudices, and misinformation. This caused a polarization of the two sides, where the authorities and vaccinated people blame non-vaccinated for their choice, and

non-vaccinated people react strongly and are even less incentivised at considering the option of getting the vaccine.

In cases like these, two solutions are possible.

- **Solution 1**: *Redefine experts, bet on competence*

The first solution deals with the first of the two reasons of failure of the vaccination campaign. This solution consists in redefining the experts who are in charge of these decisions and get to communicate them to the general public.

In this particular case, this solution would consist in highlighting the role of doctors and scientists in the decisions about vaccinations, instead of making it a pure political matter. Show that the proper experts are the scientists who know how the vaccine is made, its effectiveness, and the possible bad consequences of it. Let them communicate more these things to the public, invest on showing to the public that vaccines are safe and effective, and that there is no reason to be afraid, on the basis of scientific evidence and real discussion; a discussion that also includes the point of view of people in fear and uncertainty. Bet on the competence of the real experts and leave them in charge of the application stage.

When trying this kind of solution we do not re-enter the Genealogical Stage of creation of social kinds, since we do not want to change the constitutive rules for the kind in question.

What we do is to improve the process of application, acting on the social environment and making it more prone to adopt the kind as an operative kind in the community. The solution generalises to all cases where the application failed because of poor communication by the authorities.

- **Solution 2**: *Change strategy of application*

The second solution concerns the second reason of failure of the vaccination campaign. It consists in changing the strategies of application, rather than just replacing the people in charge of it.

Remaining within our example, one solution of this sort would be to give incentives to people who decide to get the vaccination, in order to give them a further reason to choose for it. You could give them immediate reward, like a small amount of money, of a certificate for some discount to buy what they need. But you could also give them a reward such as a discount on the next light and gas bill, or something of this sort.

Another solution would be to change the status of vaccination from discretionary to mandatory. This is exactly what is happening in Italy in the real world. In this way, instead of an incentive because of a reward, people would get an incentive due to the will of avoiding repercussions.

Generalising, we can once again see that this solution does not re-enter a Genealogical Stage but just tries to force the application of the social kind in question, without modifying its constitutive rules.

- **Case 3**: *Defectiveness due to environmental changes*

The third way for a Top-down social kinds to be defective is for it to become unfeasible due to environmental changes. New evidence, knowledge, or circumstances could make a social kind which was functional before that point to become obsolete. Once again, it is by philosophical and empirical analysis that we can be able to recognize this kind of defectiveness.

Think about the two following examples.

Consider the case of the social kind of *palm oil*. Like before, this is a good example of Top-down kinds, because there is a physical entity ‘palm oil’ that is given a social

function of being ‘food’ by some authorities, i.e. the food industries that use it as ingredient for their recipes. It is indeed rare for home cooks to use palm oil as an ingredient to cook, thereby making its use almost monopolised by these industries.

Now, in the recent years new studies and evidence have shown that palm oil is a dangerous ingredient because we now know that it is unhealthy for us. The environment has changed in this scenario, since we now have some new knowledge that renders the original kind defective. It is this environmental change that caused the kind to be flawed, and its defectiveness is now to be found in the constitutive rules, without any responsibility, neither malicious nor sincere, of the authorities in question.

Another example concerns the social group of *hysteric* women. Until the 19th century, certain physical and behavioural traits of women were believed to be symptoms of a disease, called ‘hysteria’. We are in front of a Top-down social kind because it was the doctors of the time that stipulated the term and the properties that were affiliated with the kind. Hysteric women were treated and cured in specific ways.

Then, studies in medicine and psychology revealed that there is no disease as ‘hysteria’. The symptoms that were previously connected with hysteria turned out to be better explained as consequence of depressive states, thereby making the classification of it as a disease untenable. Also, this new knowledge showed that women diagnosed with hysteria were unproperly cured and unfairly treated. Also in this case, it is due to environmental changes in our research and knowledge that the social kind became obsolete. It was not the fault of the doctors that diagnosed it as a disease since they lacked relevant evidence.

- **Solution 1**: *New Genealogy with authorities*

In cases like these, a viable solution is to enter a process of new Genealogy, in order to create a new kind that better suits the changed environment. One way to do this is to find a new group of authorities, or to use the current authorities but making sure that they are familiar with the new evidence.

For instance, in the case of *palm oil*, it is not conceivable to shut down all the industries that used it as an ingredient and create new industries from scratch. It would be much better to convert the old industries in order for them to produce food which is palm oil free. In our terms, we make sure that the authorities are aware of the new evidence, and we make them find solutions, create new constitutive rules, new recipes, and replace their unhealthy food.

On the other hand, in the case of *hysteria* we could have taught the doctors at the time the new knowledge, in order for them to be able to better diagnose the relevant conditions, or we could stop letting them diagnose that kind of disease and bring in new doctors and scientist with better awareness so that they will create the new constitutive rules and reshape the kind.

In both cases, both authorities and rules are challenged and modified, so that a new Genealogy stage is begun.

Cases like this happen also when brand new circumstances occur, not just when we acquire new evidence. For instance, new phenomena might arise, and they might require regulation and classification. Think about the kind *cybercriminal*. This is a Top-down kind because it is the law that defines who is a cybercriminal, and the law is the result of the decision and agreement between authorities. Before the second half of the 20th century, the kind *cybercriminal* obviously did not exist, since there was no cyber world for anyone to be a criminal in. New circumstances such as the diffusion of computers

and the start of bad actions performed by people with them brought up the necessity of regulations and classification. With the help of experts in the field, politicians set up new limitations to what can be done with the computer, and established *cybercrimes* and *cybercriminals*.

- **Solution 2**: *New Genealogy with public participation*

The second possible solution when environmental changes make a certain social kind obsolete and defective is to enter a new process of Genealogy by encouraging public participation. This solution is restricted to only a few instances. It is only viable for decisions that require no particular expertise and yet are relevant enough to call for everyone to give her opinion on the matter.

This is what happens, for instance, with *referendums*. Consider the Top-down social kind *water* in a hypothetical state where water has always been public and distributed by the state itself. This is a Top-down kind because on top of a physical entity 'water' there are social functions in terms of regulations and agreements, and such regulations are stipulated by the authorities. Suppose that this state is now subject to environmental changes such as the impoverishment of the national bank. To face such change, the authorities propose to sell water to private companies for them do bottle it and distribute it. The people of the state, however, are uncertain about this measure, and they call for a *referendum* to decide what to do with *water*. The public is therefore included in the Genealogy of the new kind *water*, and it participates in the stipulation of constitutive rules for it.

For kinds that are not subject to laws and regulations, however, such public participation would not be possible. We are going to discuss what regular people can do

for the modifications of kinds in the section about Bottom-up kinds. In Top-down cases, there is no room for this to happen.

- **Feasibility and control**

We have analysed three cases of defectiveness for Top-down social kinds and how to try to fix them. This is what Reality Engineering is all about, use philosophical analysis to solve practical problems about the categories at work in our societies.

As a conclusion for this chapter, I would like to address the possible concern of an enterprise like the one of Reality Engineering being unfeasible, unpracticable and uncontrollable.

These are indeed some of the accuses faced by our rival theory Conceptual Engineering, and it is important to see how this theory deals with those same problems.

About feasibility, Top-down Reality Engineering is surely a very complicated enterprise: changing things in the world is not easy, and it is not supposed to be easy. The advantage that Reality Engineering has over Conceptual Engineering is, hopefully, that the difficulties that we face are solely practical, and not theoretical.

In Conceptual Engineering, it is not too clear what it means to modify concepts, not even from a theoretical standpoint. In Reality Engineering, our philosophical analysis made very clear what can and should be changed. We have to change concrete, practical aspects of our societies, on a political, jurisdictional, and regulatory level.

Practical difficulties are to be expected. We have seen that, if we want to modify Top-down social kinds, we have to act upon authorities, experts, rules, and laws. Surely, it is not an everyday matter, say, to replace politicians and governments with new authorities, as it is not so simple to modify laws and regulations that interest our daily

lives. But we do have the necessary tools to make these changes happen, since we have the luck to live in democratic systems.

Reality Engineering does not happen with a snap of our fingers, it is a long, slow, and complicated process. The bigger the magnitude and influence of the kind that we want to change, the harder it is going to be for us to make the change happen.

The issue of Control concerns the idea that we our actions sometimes do not have the effects that we desire. In Conceptual Engineering, there is the worry that we do not really have control over the meanings of words and concepts. We therefore might not know what our attempts to change such meanings and concepts might end up doing.

In Top-down Reality Engineering, we do have control over the Genealogical stage, i.e. the process of creation of proto-kinds. Such proto-kinds are the result of the constitutive rules that the authorities stipulate and agree upon. Therefore, the authorities have complete control of the nature of the proto-kind. When we lose control is in the Application stage. As we have said, such stage is somewhat mysterious, and we never really know how a particular community and environment might react to a new kind being introduced. We have also said that there are many ways in which the Application stage can go wrong, probably more than the ones that we analysed above.

In conclusion, it is a hard process to modify authorities and rules, and we probably do not have full control on the application of the kinds that we create. I do not think that these are surprising features, nor undesirable.

Chapter 8 – Reality Engineering for Bottom-up social kinds

In the previous chapter, we have analysed projects of Reality Engineering for Top-down social kinds. We have identified the three main reasons for defectiveness of Top-down kinds and proposed solutions and actions to fix all three of such forms.

First, we discussed Top-down kinds defective due to their constitutive rules, and we recognized the possible sources of such defectiveness in the agreement between the authorities, which could have acted maliciously, incompetently, or mistakenly. We have seen two possible solutions, one where we change the authorities that are in charge of the creation of proto-kinds in the relevant framework and start a new Genealogical stage. The other solution consisted in keeping the same authorities but showing them the reasons for their mistake and pushing them to re-evaluate the rules.

Secondly, we discussed Top-down kinds that are defective due to a failure in the Application stage. This sort of defectiveness is situation specific, and there might not be generalizable reasons for the failure of application. We took the case study of *vaccine* and proposed two solutions. The first suggests to redefine who counts as expert, and the second suggests to modify the strategies for application.

Third, we discussed Top-down kinds that are defective due to environmental changes that made their constitutive rules untenable. As a solution, I suggested to enter a new process of Genealogy, thereby creating a new kind more suited for the changed environment. We also discussed the very limited cases where such Genealogy requires public participation.

In the present chapter, I will use the same strategy to analyse projects of Reality Engineering for Bottom-up social kinds. As before, we need to recall the main features

of the view and where we identified room for plasticity within it. These two steps will help greatly to clarify the case studies that I will later propose.

The analysis of Bottom-up social kinds relies on one key notion: the notion of social norms. Therefore, in this recap I will shortly present what social norms are, why external norms are metaphysically and explanatory more relevant than internal norms, and how widespread external norms have to be in order to the corresponding social kind to be operative in the relevant community.

- social norms are best understood as sets of beliefs, attitudes, practices, behaviours, and linguistic uses held and performed by people that are directed towards a particular group of other people or things *qua* being believed to be part of that group. Social norms are divided in internal, structuring, and external. Internal norms are held by people that are members of a particular group towards themselves; such norms tell people within the group as they are supposed to behave *qua* being members of the group. structuring norms are held by people that are members of a group directed towards other members of the same group: they tell them how to behave with respect to other members of the group. External norms are held by people that are not members of a group of kind directed towards people and things that are members of such groups or kinds: such norms tell them how to behave towards members of groups and kinds.
- external norms are prevalent for metaphysical and explanatory reasons. I have argued for the metaphysical sufficiency of external norms for the existence of social groups by showing that there are cases where there are no internal norms at work, but we are in front of genuine social groups. This is the case, for

instance, of *dead* and *witch*. We are in front of genuine social groups because the fact that people (or dead people) were believed to be members of these groups led them to receive specific treatment and to have a specific status, solely in virtue of them being believed to be members of these groups. I have also argued that external norms are explanatory prevalent: when we analyse the status of members of a certain social kind, it is best to look at the external norms directed towards them, because it is through external norms that rights, opportunities, and recognition are granted to them.

- How widespread external norms are is best explained by saying that for every individual in the community either she accepts the norms actively and behaves accordingly, or she behaves accordingly to it without knowing it, or she rejects it consciously. Only one option is not viable in this framework: you cannot unconsciously reject a social norm without facing consequences. There is, however, room for dissidence against such norms. But if norms are widespread in these terms, such dissidence is subject to possible negative repercussions.

About plasticity, I argued that Bottom-up social kinds are plastic, and that their ability to change over time is due to the nature of social norms. Social norms evolve and can be changed, in fact social norms change all the time in society, and such change is not always motivated by actions directly intended at making it happen. Therefore, Bottom-up kinds are intentionally-free plastic. We have also said that since social norms are a matter of beliefs, practices, and behaviours and all these things can be changed and acted upon, it is clear that intention-bound plasticity holds. In front of some kind of defectiveness in the social norms at work, we might regulate our behaviour and engage

others to do the same. If we succeed in modifying widespread behaviours and beliefs, the social norms will change accordingly.

Like in the case of Top-down kinds, also Bottom-up kinds can be defective for different reasons, due to the fact that social norms can be defective for different reasons. I will here take the three main reasons for defectiveness of social norms, analyse them, and propose solutions to them.

These three forms are the following: (1) external social norms can be harmful to members of the group, (2) social external norms can be harmful to non-members of the group/kind, and (3) social norms can be defective due to environmental changes. Let us then take on each of these separately.

- **Case 1**: *external norms harm members of the group*

The first way for a Bottom-up kind to be defective is if the external social norms attached to it harm members of the kind. Harm, however, is usually defined as a physical and/or mental damage. This means that harm can only be perpetuated against humans (and probably animals). Therefore, the first form of defectiveness only applies to social groups, and not social kinds, because you cannot harm physically or mentally inanimate things.

We are going to discuss the case of animals in Case 3, so for now I will only focus on social groups that have humans as members. In such case, there are two ways in which external social norms can lead to physical and/or mental damage to people. The first case concerns instances where external social norms are discriminatory towards a certain group: there are external norms at work and these external norms are explicitly and directly oppressive against the group. The second case concerns instances where a group fails to be recognized by the community: there are no external norms at work, and

this lack of norms causes the lack of right, opportunities, and fair treatment for members of the group. Let me analyse the two cases separately.

- **Case 1a**: *external norms discriminate members of the group*

The case of external norms discriminating members of a certain social group is the most diffused, dangerous, and serious form of defectiveness of social kinds in our society, with all due respect to the issues that we will discuss later, because they are terribly serious as well.

However, the most evident cases of social injustice at work these days are connected to this kind of defectiveness. This kind of defectiveness is the one that interests dynamics of *sexism* and *racism*.

Let us think about the social group of *women*. It is through simple empirical observation that we can see that external norms concerning women, i.e. the practices, beliefs, behaviours that men hold with respect to women, are oppressive towards them. Women are too often harmed physically and mentally harmed by their partners, denied job occasions with no reason except for their being a woman, mistreated by their colleagues and peers, and so on. I cannot here enter in the discussion of women oppression, but the point is just that we have all the theoretical and empirical evidence to conclude that women are unfairly discriminated in our societies, and that this has to stop.

Similar horrible dynamics are in place against social groups associate with race, such as the social group of *black* people. Also in this case, we can observe constant instances of mistreatment, denial of rights, denial of equal opportunities, abuse coming from colleagues, superiors, police forces, and so on. Again, I cannot enter a detailed discussion about this because it would lead us off point. In this framework, the key

point is that social norms directed towards *black* people are obviously harmful, physically, and mentally. This is enough for us to say that these norms are defective.

But why do we have these norms? Why are they so widespread if they are so clearly flawed?

The answer is not straightforward at all. Social norms such as those concerning *women* and *black* people come from a mixture of historical, cultural, and religious factors, at least in western societies. Again, I want to stress that I am aware of the oversimplification that I am guilty of. I hope to convey that my point is not to provide a detailed analysis of these case studies but to understand the general ways in which social groups and kinds are defective and how to go about this defectiveness.

In the paragraph about solutions for this kind of defectiveness I will argue that where the norms come from is important but not fundamental. If our sensibility today is changed, i.e. if we now want to ground our behaviours on principles of human dignity, equality, and parity of rights, it does not matter if we behaved unfairly for thousands of years. The point is to change now.

- **Case 1b**: *external norms fail to recognize members of the group*

Another important way for social norms to be defective is when they fail to recognize people that identify themselves as members of the same group and act according to internal norms as genuine social groups.

This is the case, for instance of people that identify themselves as *non-binary*. *Non-binary* people usually self-identify as not suiting the ordinary twofold division of genders. They identify as not entirely male nor entirely female. I say nothing uncontroversial by noting that members of this not recognized group face sever difficulties in their social lives.

In this framework, we can explain the source of such social difficulties as caused by a failed recognition of this group of people as a genuine social group. This means that there are no external norms in place, and that people behave towards member of this group in diffident, misinformed ways. *Non-binary* people are not yet a genuine social group in the way that I defined, but they should be.

Due to complicated reasons that have to do with history, culture, religion, fear, and prejudice we are failing to accept the social group of *non-binaries* as an active component of our societies.

- **Solution 1a**: *de vetero Engineering*

The first case of defectiveness that we have seen concerns social norms that harm members of certain social groups, these norms being discriminatory, oppressive, or dangerous in other ways.

In such case, we have an operative social group which is flawed, and we want to modify it, in contrast with Case 1b where we do not have an operative social group, but we want to introduce it. Therefore, Case 1a is in need of *de vetero* Reality Engineering, and Case 1b is in need of *de novo* Reality Engineering. This *de vetero*/*de novo* distinction has been introduced by Chalmers in Conceptual Engineering (Chalmers, forthcoming), to indicate that conceptual introduction and conceptual re-engineering are two relevantly different projects. I think that the same goes for Reality Engineering: kind introduction and kind revision are relevantly different and worth analysing separately.

De vetero engineering for Bottom-up social groups, which is the solution for Case 1a, aims at debunking old defective norms and replace them with better new ones. The project is complicated, since we have said that external norms ground social groups

only when they are substantially widespread. Therefore, in order to change such norms, we have to able to persuade people of their defectiveness.

I propose a three-step process for this project: *transparentise-expose-propose*.

- *Transparentise:* the first step is to understand where the defective norms come from. This is a theoretical enterprise, which involves historians, philosophers, anthropologists, psychologists, and sociologists. This process is extremely important, because understanding the origin would help understanding what beliefs, behaviours and practices of the past led us to have the norms that we have. The importance of this step is to facilitate the next one.
- *Expose:* this step is the crucial one. In order to persuade people of the defectiveness of their beliefs and behaviours, we have to expose *that* such beliefs and behaviours harm others and make them question *why* they are maintaining such behaviours and beliefs. In the case of *women*, on a small scale we can show that bad individual attitudes like catcalling or body shaming or (even worse) physical and mental abuse are harmful and cause dangerous consequences to the victims. Many of the perpetuators of these behaviours think that they are not causing harm, that their behaviour is normal and pacific, and this is the first thing that we have to change. Then, we also have to expose the reasons why people behave in the way they do. It is important that, even if we do not perfectly know the historical/cultural origin of bad norms, we show that they are grounded on flawed principles, that are contrary to our contemporary morality and sensibility. We must lead people to see that their practices harm others and make them question why they are doing such things. On a large scale,

we can also try to act on the discriminatory policies that are in place, working to make sure that people get similar treatment, rights, occasions, and retributions.

- *Propose:* the last step represents the constructive part of Reality Engineering projects. If we want to modify bad social norms, we have to create new ones. We have to start acting differently on an individual level and campaign for these new behaviours to be diffused. We have to imagine new ways to relate with other people and work for these new ways to take place. Bottom-up Reality Engineering heavily relies on personal responsibility because beliefs and behaviours are, ultimately, a matter of individuals.

- **Solution 1b**: *de novo Engineering*

The second case of defectiveness that we have seen concerning social norms and Bottom-up kinds is when we have a lack of norms concerning a particular group of individuals, and this lack of recognition causes harm.

In such case, we do not have an operative social kind, and we would want to introduce it. Therefore, we need a project of *de novo* Reality Engineering. We have said repeatedly that Bottom-up social kinds are grounded in social external norms. Therefore, if we want to create a new social kind from scratch, we have to create new social external norms.

For this, the most important part of the job is for the members of the non-recognized group that identify themselves as members of the group and share internal norms. Since their problem is lack of recognition, their top priority is to gain some visibility as genuine and proper members of our society. For instance, if everyone really knew who are *non-binary* people, what they think and why they identify in a certain way, at the

very least they would form beliefs towards this group, and these beliefs can be the starting point for external norms. The risk is that people could form negative beliefs, thereby creating bad external norms. But in that case a social kind is at work, and we could work with it via Solution 1a.

The most difficult case is when people that should be considered part of social groups that should be given more rights and protection do not and cannot share internal norms. This is the case of people with special needs and/or disabilities that are not enough safeguarded. This is an extreme case, because there are no external norms in place (since people may not be aware of the special needs of others) and no shared internal norms in place (since people might not know that they have special needs or might be unable to communicate it). In such cases, our hope relies on sciences to seek these issues, and once science produces the relevant natural kind, we can form a social kind that helps identifying and protecting these individuals.

- **Case 2**: *external norms harm non-members of the group/kind*

Another case of defectiveness for Bottom-up social kinds (or groups) happens when external norms harm non-members of the group/kind. This means that the beliefs and behaviours that people have towards a kind harm those that hold them.

Of course, this happens only when people that hold such beliefs and behaviours are not aware of the harm that they causing to themselves, usually because such damage is not directly experienced by them, but only indirectly, or unconsciously.

To clarify, imagine a community that supports that social group of *dictators*, by trusting whoever is a member of such group and never disagreeing with them. In exchange, they are forced to live in a brutal system, where they are deprived of rights and of what they need to prosper. These situations are not uncommon, especially in places where people

never saw what a democratic society looks like, and they are unaware of their possibility of creating a better community.

Also in our societies, however, we have instances of this sort of defectiveness. Think about the social kind *smartphone*. Most people own a smartphone, use it, have attitudes and beliefs towards it. In other words, we have external norms towards the kind s*martphone*. Most of us look at smartphones as extremely useful tools, to communicate, read newspapers, listen to music, and so on. However, smartphones can harm those who use them, when misused or overused, in different ways: they reduce our privacy, they damage our attention span, they damage our ability to communicate and relate to others, and so on. It can be argued that our external norms towards smartphones damage us.

- **Solution:** *Transparentize, people do not want to harm themselves*

Against this form of defectiveness, the solution is theoretically pretty easy, and it is to transparentize the norms that we have and their negative effects.

By transparentizing the norms and informing those who hold them, we make them aware of the fact that their own beliefs and behaviours harm them. If we manage to provoke in someone the belief that what he is doing is actually harming her, that person is extremely likely to stop/reduce what she is doing.

However, there might be practical difficulties. If in the case of smartphones a belief that overuse really harms you is likely to cause you to reduce its use, cases such as the one of *dictators* are not so straightforward. If the people of our hypothetical community stopped to support the dictators, they would be likely to be harmed even more, at least in the initial stage of their 'rebellion'. In cases like these, conflict is usually unavoidable, and Reality Engineering happens as a slow and painful process.

- **Case 3**: *Defectiveness due to environmental changes*

The third kind of defectiveness for Bottom-up social kinds concerns the case of external norms being obsolete. This happens when norms that were acceptable in a certain circumstance become untenable due to a change of circumstances, i.e. a change in the social environment where they are at work.

This is similar to the third way of defectiveness of Top-down social kinds, and the reasons that ground such defectiveness are similar as well.

As before, new evidence, knowledge, or circumstances could make a social kind which was functional before that point to become obsolete. Once again, it is by philosophical and empirical analysis that we can be able to recognize this kind of defectiveness.

Think about the social group of *vegans*. The external norms that people have towards members of this group lead them to be diffident, and sometimes disrespectful. But this is not the key point of the example, because this kind of treatment would make the defectiveness of the social group of *vegans* fall under Case 1.

The point is that these social norms prevent most people from identifying as *vegans* and behaving in accordance with the internal norms that characterize such group, which is to not eat, use or consume products that are derivate from animals.

Up until at most one century ago, there were perfectly good reasons for not being *vegan*, reasons so compelling that the group of *vegans* did not exist at all. The reasons were that most people needed to consume animal products in order to survive and thrive. This motivation trumped all moral concerns towards animals and the environmental concerns about sustainability (which was not a huge problem until half a century ago).

We can say that even if eating animals has always been a moral and environmental problem, such problem was worth the gain that we got from such behaviour: our survival.

Today, however, the social and natural environment changed dramatically. We now have compelling scientific evidence that concludes that plant-based diets are perfectly healthy, and we do not need to eat animals in order to survive and thrive (Craig, Mangels, & Association, 2009). We have the technology and capacities to produce enough variety of plant-based foods to cover all our nutritional needs.

Therefore, the main reason why eating animals was morally and environmentally acceptable is no longer applicable[8]. In other words, our social norms towards *vegans*, and especially our refusal of their internal norms is defective.

- **Solution**: *Transparentize, adopt internal norms, campaign*

As is the solution for Case 2, what to do in these cases to fix defectiveness is theoretically straightforward. Since the most rational and moral solution is to change the external norms that we have, transparentising the bad effects of these norms is the first move to make.

By showing the moral and environmental concerns connected to eating animals, for instance, we can at least make people question their beliefs and behaviours.

The second step concerns individual responsibility. If we believe that certain external norms are mistaken and that we should change them into new and different internal norms, the correct move is to adopt those internal norms ourselves. Once again, large scale modifications in beliefs and behaviours can only happen with small scale changes. We have to be the living examples of the social norms that we want in our societies.

Third, if we believe that these new social norms are important and worth sharing, we should campaign for them. We should work for them being more transparent and widespread, and this can be done in different ways. The more widespread the norms

become, the stronger the social group will be, the easier it will be for these norms to finally be established in our communities.

- **Feasibility and Control**

The discussion of problems of Feasibility and Control for projects of Reality Engineering for Bottom-up social kinds does not particularly differ from the one that we had about Top-down kinds.

In our analysis of cases and solutions, we have seen that the main theoretical difficulty of these project is the understanding of the origin of certain social norms. Such understanding, as we have said, would benefit the process of transparentising the norms, which would help showing people of the effects that these norms have, and why we have them.

The project, however, remains theoretically feasible: I argued that even if we fail to recognize the origin of defective social norms, we can still expose their problems. In this case, the difficulties will be practical.

About practical feasibility, the point is the same that with Top-down kinds: engineering social entities is not easy, and it is not supposed to be so.

The advantage of Bottom-up kinds, however, is that personal responsibility and individual behaviours are fundamental. Therefore, even if we cannot modify social kinds just by acting differently, our individual attitudes relevantly matter to the enterprise. With Top-down kinds, we did not have such an important role, given that the constitution and application of kinds came from authorities.

Control is a complicated issue. We definitely have control over our own actions and beliefs, but we cannot know a priori what kind of beliefs and behaviours people will adopt once the original social norms are undermined. A much richer analysis would be

required to answer this point, but not a philosophical one. We should look for help to psychology and sociology to tell us what these major behavioural changes could/would look like.

The conclusion is similar to before with Top-down kinds: it is a hard process to modify people's beliefs and behaviours on a large scale, and we probably do not have full control on the twists and turns that these modifications can take, thereby not being certain that the new external norms and the corresponding social kinds will end up being operative in the way that we desire. I do not think that these are surprising features, nor undesirable.

Conclusions

To conclude this dissertation, I would like to briefly highlight the key points and arguments made. Then, I will provide my opinion on the significance of this project in the Conceptual Engineering debate and in metaphilosophical discussions.

I started off by presenting Conceptual Engineering as a philosophical and metaphilosophical movement. Its main goal is to find defective concepts and fix their defectiveness.

Some Conceptual Engineers even argue that by modifying the meaning of a word (or a concept), we also get to change the nature of the thing in the world which is referred by that word (or concept).

Following Greenough, I criticized this worldly characterization of Conceptual Engineering. I presented Greenough's objection about the ontological baggage of the view, which is committed to a weird linguistic structuring of the world, and I posed an original challenge to the view: the Insufficiency Challenge concerning social Concepts and social Entities.

These two reasons motivated my jump towards Reality Engineering, which is a more practice-oriented movement, where we focus directly of changing things in the world.

Then we discussed the issue of Scope for Reality Engineering and argued for kinds over properties as the subject matter of the discipline.

Afterwards, I scouted four social Constructivist views on the metaphysics of social kinds: the Constitution Account, the Conferral Account, the Entrenchment Account, and the Normative Approach.

Out of these four views, I extrapolated two general theories on social kinds: the Top-down pattern and the Bottom-up pattern. According to the Top-down pattern, social kinds are created via stipulation by some authorities, in terms of constitutive rules, and they are applied on society. According to the Bottom-up pattern, social kinds emerge from social external norms that are widespread in the relevant community.

Then, I used some case studies to show how we can conduct Reality Engineering projects with respect to both Top-down and Bottom-up kinds. When doing this, I also discussed the practical difficulties that are to be expected in trying to modifying social kinds.

The significance of the project, in my opinion, lies in its extreme practice-oriented turn. Conceptual Engineering represented a good starting point towards the practical applications of philosophy. However, as we have seen, Conceptual Engineering is too niche and theoretical: we are not going to change the world by changing the words that we use!

Reality Engineering aims to use philosophy to change things for the better. Philosophical analysis is here just the starting point of the project of modification, and not the end of it.

In order to accomplish her goals, a Reality Engineer has to stand up from her philosophical chair and actively engage with the world, on an individual, communitarian and political level.

Is philosophy ready to gambit its theoretical golden armour in order to serve the change?

Bibliography

Ásta, K. S. (2013). The social Construction of Human kinds. *Hypatia, 29 (4)*, 716-732.

Belleri, D. (2020). Minimal Substantivity. In A. Burgess, H. Cappelen, & D. Plunkett, *Conceptual Engineering and Conceptual Ethics* (pp. 59-78). Oxford University Press.

Boghossian, P. (2006). *Fear of Knowledge: Against Relativism and constructivism.* Oxford University Press.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. Wilson, *New Interdisciplinary Essays* (pp. 141-185). MIT University Press.

Cappelen, H. (2018). *Fixing Language.* Oxford University Press.

Chalmers, D. (forthcoming). What is Conceptual Engineering and What Should it Be? *Inquiry: An Interdisciplinary Journal of Philosophy*.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis 58:1*, 7-19.

Craig, W., Mangels, A. R., & Association, A. D. (2009). Position of the American Dietetic Association: vegetarian diets. *Journal of the American Dietetic Association 109 (7)*, 1266-82.

Diaz-Leon, E. (forthcoming). Substantive metaphysical debates about gender and race: Verbal disputes and metaphysical deflationism. *Journal of social Philosophy*.

Greenough, P. (2019). Conceptual Marxism and Truth: Inquiry Symposium on Kevin Scharp's Replacing Truth. *Inquiry 62:4*, 403-421.

Greenough, P. (forthcoming). Conceptual Engineering via Reality Engineering.

Gustavo Isaac, M. (2021). Post-Truth Conceptual Engineering. *Inquiry: An Interdisciplinary Journal of Philosophy*, 1-17.

Hacking, I. (1999). *The social Construction of What?* Harvard University Press.

Haslanger, S. (2000). Gender and Race: (What) Are They? What Do We Want Them to Be? *Noûs 34:1*, 31-55.

Haslanger, S. (2005). What are we talking about? The Semantics and Politics of social kinds. *Hypatia 20:4*, 10-26.

Haslanger, S., & Saul, J. (2006). Philosophical Analysis and social kinds. *Proceedings of the Aristotelian Society, Supplementary Volumes (80)*, 89-143.

Hirsch, E. (1993). *Dividing Reality.* Oxford University Press.

IPCC (2019). Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems [P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, (eds.)]. In press.

Kaspers, T., LiBrizzi, J., Calosi, D., & Kobe, Y. (under review). Speak no ill of the dead: The Dead as a social group .

Kerr, E. T. (2014). Engineering Differences between natural, social and artificial kinds. In M. Frassen, P. Kroes, T. A. Reydon, & P. E. Vermass, *Artefact kinds* (pp. 207-225). Springer.

Kusch, M. (1999). *Psychological Kowledge: A social history and philosophy.* Oxford: Routledge.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy 61 (4)*, 343-377.

Mallon, R. (2016). *The Construction of Human kinds.* Oxford University Press.

Marques, T., & Wikfross, A. M. (2020). *Shifting Concepts: The Philosophy and Psychology of Conceptual Variability.* Oxford University Press.

Plunkett, D. (2015). Which Concepts Should We Use?: Metalinguistic Negotiations and The Methodology of Philosophy. *Inquiry: An Interdisciplinary Journal of Philosophy 58 (7-8)*, 828-874.

Plunkett, D., & Sundell, T. (2013). Disagreement and the Semantics of normative and Evaluative Terms. *Philosophers' Imprint, 13 (23)*, 1-37.

Podosky, P.-M. C. (2018). Ideology and normativity: constraints on conceptual engineering. *Inquiry: An Interdisciplinary Journal of Philosophy*, 1-15.

Ritchie, K. (2015). The metaphysics of social groups. *Philosophy Compass; 10 (5)*, 310-321.

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

Scharp, K. (2020). Philosophy as the study of defective concepts. In A. Burgess, H. Cappelen, & D. Plunkett, *Conceptual Engineering and Conceptual Ethics.* Oxford University Press.

Searle, J. R. (1995). *The Contruction of social Reality.* free Press.

Simion, M. (2018). The 'should' in conceptual engineering. *Inquiry 61:8,* 914-928.

Tanswell, F. S. (2018). Conceptual Engineering for Mathematical Concepts. *Inquiry: An Interdisciplinary Journal of Philosophy 61 (8)*, 881-913.

Thomasson, A. (2003 (a)). Foundations for a social ontology. *ProtoSociology 18*, 269-290.

Thomasson, A. (2003 (b)). Realism and Human kinds. *Philosophy and Phenomenological Research, 67 (3)*, 580-609.

Thomasson, A. (2015). *ontology made easy*. Oxford University Press.

Thomasson, A. (2019). The ontology of social groups. *Synthese; 196*, 4829-4845.

Thomasson, A. (2020). A Pragmatic Method for normative Conceptual Work. In A. Burgess, H. Cappelen, & D. Plunkett, *Conceptual Engineering and Conceptual Ethics* (pp. 435-458). Oxford University Press.

Tsou, J. (2020). Social Construction, HPC kinds, and the Projectability of Human Categories. *Philsophy of the social Sciences, 50 (2)*, 115-137.

Wittig, M. (1992). *The Straight Mind and Other Essays.* Boston: Beacon Press.

[1] There is some disagreement in the literature about the labelling of the subject matter. Ásta and Mallon use "Human kinds", Thomasson used "institutional kinds" in 2003 and "social groups" in 2019. I do not think that much attention should be dedicated to this detail. I will use the label "social kinds" to accommodate for all of these alternatives.

[2] This view only accounts with categories that include persons as their members. So, by restricting the focus to "Human kinds", Mallon wants to avoid discussions about kinds like *money*. I do not see, however, any complication in the application of his account to categories that include also non-human entities as their members. Therefore, I will keep talking about social kinds, implicitly including these latter categories as well.

[3] A reader familiar with Ian Hacking's theory of the *looping effects* of Human kinds might recognize some similarities with this discussion.

[4] Such anti-metaphysical approach really separates Thomasson's position from 2019 to the ones she held in 2003. As testimony of such separation, it is interesting to notice that, despite the similarity of topic, Thomasson does not include either of her papers from 2003 as references for the 2019 paper.

[5] This view only accounts with categories that include persons as their members. So, by restricting the focus to "Human kinds", Mallon wants to avoid discussions about kinds like *money*. I do not see, however, any complication in the application of his account to categories that include also non-human entities as their members. Therefore, I will keep talking about social kinds, implicitly including these latter categories as well.

[6] We might have complications in the case of Animal kinds, members of which might indeed share some internal norms. However, at this point, I am inclined to think that Animal kinds are part of the realm of natural kinds rather than in the one of social kinds.

[7] I do not believe that there are relevant differences between *replacement* and *revision* in Conceptual Engineering, I think it is a matter of philosophical taste whether you think that you are creating a brand-new concept or simply modifying the old one. The point made here is just to better illustrate the dynamics at work.

[8] I cannot enter detailed moral discussions about veganism, nor I can provide too many details on the environmental impact of eating animals. If interested in the latter, please check the IPCC Special Report on Climate Change that I put in my bibliography.