THESIS

DOING BETTER THAN TRUTH

THE CONCEPTUAL ENGINEERING OF A BASIC CONCEPT

Submitted by

Jordan Davis

Department of Philosophy

In partial fulfillment of the requirements

For the Degree of Master of Arts

Colorado State University

Fort Collins, Colorado

Spring 2025

Master's Committee:

Advisor: Jeff Kasser

Collin Rice
Ben Prytherch

ABSTRACT

DOING BETTER THAN TRUTH:

THE CONCEPTUAL ENGINEERING OF A BASIC CONCEPT

Truth can and should be replaced as a concept, or at least we should strongly consider doing so. Conceptual engineering is about determining what concepts we should use and the modification and creation of concepts to serve that purpose. One of the ongoing areas of research is determining just what the limits of conceptual engineering are. I approach this topic by exploring the possibility of conceptually engineering truth by replacing it as a concept and what that possibility means for replacing other concepts. More specifically, I use Kevin Scharp's proposal for replacing truth as model for how and why we might replace truth and how that might generalize to replacing other concepts. After a detailed discussion of the mechanics and motivations of Scharp's proposal, I argue that any substantive distinction between replacement and the revision fails because of the messiness of conceptual identity, and that, consequently, replacement is pervasive in conceptual engineering and philosophy more broadly. I continue by exploring various objections against the possibility and permissibility of replacement, focusing on truth in particular. I then show that under a framework where we treat concepts as tools that fulfill certain roles in addressing problems that those objections either fail or are defanged. I argue that such a framework is plausible based on how it approximates how we already think about things like concepts and the numerous benefits of adopting the framework. I explain how the framework addresses each objection both with respect to truth and basic concepts generally.

ACKNOWLEDGMENTS

As the culmination of a seventeen-year academic journey, this thesis represents the efforts of many people to whom I owe more gratitude than I could ever reasonably express here. Learning to be selective and to limit scope is a skill one must acquire early in academia. I am not sure I have yet mastered it, but I will do my best.

My advisor, Jeff Kasser, is the obvious first choice. He has had saint-like patience with me as I meandered through this project. He stuck it out with a chaos muppet like me, and for that alone, I will be eternally grateful. His comments, advice, and encouragement have been priceless—far more than I deserve.

When Collin Rice agreed to be on my committee, I had not yet had the opportunity to meet him. He accepted the invitation despite my being a total stranger. His goodwill and sense of community made my graduation and the completion of this thesis possible. His comments have been thought-provoking and have pushed me to more deeply consider certain issues. He will always have my gratitude.

Ben Prytherch was kind enough to volunteer his time and energy, going above and beyond his obligations, to serve as my outside reader. I have little doubt that being an outside reader is an almost entirely thankless job. The least I can do is ensure that he walks away with at least my thanks.

My undergraduate professors were also key to getting me to this point. Without them I would never had a chance to write this thesis. Mark Silcox taught me more than probably than other person. Eva Dadlez made me a better writer. James Mock introduced me to pragmatism,

which has made me the thinker I am. Aaron Fortune really built on that. Rick Chew cultivated my love for logic.

My cohort, as well as those before and after us, were more consistently amazing than anyone could have ever hoped for. Jackson and I made a great team from day one—I appreciate his support and his patience in putting up with me for three years as his roommate. I have to thank Steve for helping me come up with the title among other things. Commiserating with Jackie will always hold a special place in my heart. Playfully duking it out with Jesse or John over some epistemic or moral minutiae helped keep me sane. The cohort before mine made me feel very welcome—Bo, Alec, and Jacob in particular. I will always appreciate Taylor for listening to my rants and for the amazing zucchini relish. I appreciate Ella for always being down to hang. I want to thank Jed and Stephanie for, among other things, making our tabletop campaign possible. Jed was a great GM, and Stephanie was a great host. I know I am not the only one for whom the joy of gaming around the table brought a little more light to our lives.

I have deep gratitude for my loving partner, Julie, who has been my rock over the past few years. I want to thank my parents, Jill and Scott Davis—without their love and support, this journey would have been much more difficult. I am also grateful to my sister, Jessica, who has had ceaseless faith in me the whole way through. And finally, I am thankful for my long-departed brother, Christian Davis, whose long, exploratory conversations set me on this path. I also need to thank all the friends I made along my journey. Jacob and Paden have amazing friends and roommates in the long time we've known each other. My friend Josh is no longer with us, but so many of our conversations pushed me to explore in a way that I wouldn't have without them. My time with all three of them made me who I am. I will cherish those times for the rest of my days. Faris has been a great friend and vital part of who I have become. Chelsea's

impact on my life cannot be expressed. There are so many people to name that I could go on forever. But thank you to all those I could not name but who will always be woven into the fiber of my being.

DEDICATION

This work is dedicated to the memory of Christian Davis as well as the friends, family, and the fellow travelers that brought me to where I am...

TABLE OF CONTENTS

The Introduction, The Scharpian Proposal, and Replacement

0.0.0. Introduction

At bottom, this essay is a defense of the possibility and permissibility of engineering our basic concepts. In other words, I defend the notion that basic concepts[1] like *truth*[2], *reason*, *good*, and the like are appropriate targets of what has come to be called "conceptual engineering." Conceptual engineering is about improving our conceptual repertoire by either inventing wholly new concepts or replacing our current concepts with better ones. As Michael Prinzing puts it, conceptual engineering "is not about the concepts we do have, but the concepts we should have." (2018, 854) What sets conceptual engineering apart from previous methodologies in philosophy is the self-conscious prescriptivity and the goal of creating new concepts[3].

Before getting deeper into the goals of this thesis, it is worth addressing a few things. One of the things this thesis is not about is metaphysics or ontology except where it is pertinent to

---

[1]One might wonder what makes a concept count as basic. It might be that only concepts that give us abilities to perceive, something like innate concepts. This is not quite the sense of "basic" that I mean. By "basic," I mean those concepts that seem to lie near the bottom of discourse, the sort of concepts that make discourse possible as we know it. Arguably, we need *truth* or a *truth*-like concept to contrast with lies and other falsehood and as a fundamental component of inquiry. Discourse would, minimally, look quite different without truth. Given that some of that might be contestable, we can at say with confidence that these concepts are basic for philosophy as it has been practiced in the Anglophone world since at least the beginning of the 20th century.

[2] I will use italics and quotation marks to distinguish between the sort of things that a concept is supposed to capture, the word used for a concept, and concepts themselves. More specifically, I use italics when speaking unambiguously about a specific concept as distinct from its non-conceptual counterpart. I use quotations to talk about the words associated with the concepts. And I leave the text plain when discussing the supposed metaphysical entity captured by some concept and when predicating something with the associated property or in cases where the distinction between the concept or its counterpart is ambiguous or irrelevant. For example, *truth* is the concept of *truth*. "Truth" is the English word that we use to talk about both the concept of *truth* and truth. Truth represents the metaphysical entity that might exist apart from our thinking and talking about it, and there are some sentences that are true.

[3] Conceptual novelty with respect to identity is addressed in 1.3.1, i.e. whether a concept is a modification of another concept or a wholly new concept.

whether we can or should purposely change concepts. I take no position on what concepts are (mental representations, abilities, abstract objects, all of the above, etc.) except that whatever they are we can change which ones we use. For sake of maintaining an appropriate scope, I should be read as only accepting notions of concepts where such change is minimally an open question.

Another issue worth addressing is what I have in mind when speaking about replacement. I use "replacement" broadly. A new concept replaces an old concept anytime the new concept starts being used by a substantial part of a community in place of the old concept at least under some repeating context or contexts. They have replaced the old concept in those contexts. Of course, replacement might be a lot stronger than that articulation might suggest. An old concept might be replaced by enough communities in enough contexts such that it becomes the dominant concept in some society. We could call that strong replacement.

A case could be made that *binary gender* (man-woman) was at least temporarily replaced in the strong sense by *continuous gender* (degrees of man, woman, or even gender itself). Replacement, even in the strong case, does not entail to complete dropping of the old concept. Replacement is perfectly compatible with different concepts being used for something like the same purpose within a community or a society.

Ultimately, I will argue that replacing *truth* is a feasible and recommendable conceptual engineering project. For the introduction, I just want to give an intuitive sketch of what I mean by "feasible" and "recommendable." I will expand on feasibility and recommendability in sections 2.1.0 and 2.2.0, respectively. Conceptual engineering begins with proposing an alteration of an existing concept or the invention of a new concept or concepts that are intended

to be used in place of the targeted concept. There are two ways to think about feasibility as it regards conceptual replacement: theoretical and practical.

Theoretical plausibility comes down to whether some conceptual project could be done in principle. To establish whether something is theoretically feasible, we only need to show that there are no strong in-principle reasons to think that such a project could not be accomplished. By a strong in-principle reason, I just mean a reason for believing that something is or is not possible in principle, and that said reason is strong relative to countervailing reasons in some discursive context.

Some readers might wonder what could amount to an in-principle reason against replacing *truth* or other conceptual engineering. Some will take it to be obvious that replacing truth or the conceptual engineering of at least some other concepts is impossible. For those amenable, I give a preview of some of the in-principle reasons against both feasibility and recommendability below when introducing Chapter Two. For the skeptics, section 1.3.0. subsection 1.3.2. in particular, should at least give conceptual engineering and replacement an air of plausibility.

Something being practically feasible depends on whether there are concrete[4] and particular reasons why a conceptual engineering project could not be accomplished, e.g. a concept too complex to be functional for humans or a concept whose nuance is such that it will necessarily be flattened by the public into being equivalent to an already popular concept. These reasons are more like the kind of reasons to think that some social engineering projects may or

[4] Concrete might be a strange word when talking about abstract objects like concepts. Concrete here is supposed to indicate concrete situations i.e. how a concept gets employed and the dynamics of that employment with respect to living, breathing human beings and all the biological, psychological, and sociological factors at play.

may not be accomplishable. Whether something is practically feasible is primarily an empirical matter. That said, this thesis will be primarily concerned with theoretical feasibility.

A conceptual engineering project is recommendable if it has a *prima facie* deliberative considerability. When we encounter some problem that seems to originate in some concept, if we should at least strongly consider replacement as part of our initial deliberations, then replacement is recommendable. We can think of different actions or moves in games as feasible or recommendable. In the game of basketball, one can successfully make a clean pass to an opponent. Neither the rules nor human limitations prevent one from doing that. It is feasible. It is not a recommendable move, given that almost any other legal move would be better given the goal of basketball, as the goal of basketball is to have people on your team put the basketball through the correct hoop more often than the other team and giving the ball to an opponent is counter-productive to that goal.

In baseball, throwing a 115-mph pitch is recommendable insofar as if one could do it, there would be times one should, but it is not recommendable insofar as it is not feasible. We can distinguish between hypothetical recommendability and substantive recommendability, where the former does not require feasibility and the latter does. The distinction is just to show that recommendability can come apart from feasibility in some sense. Bunting in baseball is both a feasible and recommendable move in baseball. That is not to say that bunting is the best move in all situations, but that there are times and situations where it is.

A better analogy might be something like heart replacement. We can imagine someone having heart issues. We can also imagine that we might one day have artificial hearts that we can use to replace faulty human hearts. Initially, heart replacement is not possible. It could be that heart replacements remain uninvented, or it could be that we cannot reliably perform heart

replacement surgery. Eventually, heart replacements come to exist, and we can perform them. However, the quality of the heart replacements or the risk of the surgery makes the consideration of heart replacement only permissible for the most dire of cases.

At that point, heart replacements would be feasible but not recommendable. Eventually, heart replacements are about as suitable as healthy natural hearts, and surgery has become relatively safe. At this point, heart replacement is not only feasible but recommendable. Recommendable in that heart replacement is almost always among the first options considered when considering severe heart issues: it becomes part of the standard toolkit.

Something worth noting here is what sort of things might ground recommendability in a particular case. As in, what kind of normative reasons or values would motivate replacement. Correcting conceptual defects might be one such motivation: extension being too narrow or too wide, inconsistency, paradox-generation, conceptual confusion, etc. We could put this under theoretical functionality, a theoretical value or virtue among others such as parsimony, explanatory power, unification, predictive power to name a few. Moral or political values or ends might make some conceptual change recommendable.

Haslanger has argued that changing the concept invoked by "woman" will help fight injustice (Haslanger 2002, 36) for example. These can ground the recommendability for replacing some concept with another. In other words, reasons along these lines are part of what motivates consideration of some concept or concepts in particular when deliberating how to address some intellectual problem. It also along these lines that we might compare different contenders as replacements. One contender might both solve certain conceptual defects and provide more explanatory power compared to some concept that only solves the defects.

We can think of the conceptual engineering and replacement of *truth* as a test case. I will argue that establishing the feasibility and recommendability of replacing *truth* gives us good reason to think replacing other basic concepts is similarly feasible and recommendable. I want to show that replacement is always an option, if not always the best option in every case. I want to show that if we were to replace *truth*, we could replace that replacement[5]. If I can successfully establish the two prior claims, I want to argue that they together provide substantive evidence that other basic concepts could be similarly replaced.

The way that I aim to show this is to focus on replacing *truth* in response to a particular problem, and so I will defend replacing *truth* as a concept as a broad strategy for solving the liar paradox[6]. While I take the view expressed in Kevin Scharp's *Replacing Truth* as an exemplar of the strategy and a strong candidate, I take myself to be defending replacement more generally. Where this thesis goes beyond Scharp is by looking at objections to replacing *truth*, how those objections might apply more generally, and putting forward a plausible theory of purposive conceptual change that looks to answer these objections. Consequently, I aim to frame the objections presented and my responses to them as working for replacements other than Scharp's.

The thesis is broken down into three chapters. In the first chapter, I present Scharp's argument for replacement and his candidates for replacement concepts as a paradigmatic case of the more general strategy I am defending. I begin the chapter by explaining Scharp's project and

[5] It has been pointed out to me that there is a temporal component here. We might reasonably ask how fast after one replacement; we can replace the replacement. My view would be that one does have to weigh conceptual stability and human patience if we were talking about concerted efforts. Luckily, I think that these things are going to be pretty slow and organic generally, and that by the time the relevant communities adopt the replacement to the point of dominance, the concept will be worked out well enough to last a long time. On the rare exception, we might see the beginnings of another replacement begin its slow march to dominance soon after. Barring some kind of catastrophic entailment or something that gets missed, it will be slow.

[6] While for reasons of space I neglect to go into this, it does also resolve Yablo's paradox.

the problem he takes himself to be addressing before going into the mechanics of his proposal, replacing *truth* with *descending truth* and *ascending truth*. His intention is not completely replace our use "truth" and "true" or even *truth* but providing concepts that will resolve the issues that emerge when *truth* goes awry, when *truth*'s inconsistency presents itself[7].

What is going to ground the Scharpian proposal's recommendability is the solving the conceptual defect of *truth*'s inconsistency and the paradoxes that it generates, including the so called revenge paradoxes. *Descending truth* and *ascending truth* could be said to be more theoretical functional than truth, lacking the same defects as *truth*. I further explain that his project is in part motivated by the role it plays in truth-conditional semantics. Then I turn to address whether a revision/replacement distinction will make any normative difference when considering replacement.

In conversations on this topic, many have suggested that revision is inherently preferable to replacement or even that revision is possible and replacement is impossible. Given that people have these intuitions, I will show that these intuitions are unfounded before turning to my main argument that revisions should just be treated as replacements. I argue that any sufficiently strong distinction is going to be difficult to make mostly due the issues surrounding the identity of concepts. Assuming for the sake of argument that some strong distinction could be drawn, the best claim to preferability for revision is conservation. Conservation is just shorthand for lesser resource-intensiveness and less risk and those are not decisive values. Having cast doubt on there being anything distinctive about revisions, I argue that revisions are just replacements in the relevant sense. I then turn to showing that replacement is pervasive in philosophy by showing

[7] Scharp goes over this in chapter 9 in (2013).

how the notion of "truth" that many students have is often replaced by the *truth* philosophers are more familiar with.

In the second chapter, I present several objections to replacement. As I mentioned before, here I will give a more extensive preview of Chapter Two relative to the other chapters to help explain the motivation of the thesis. We start with three objections against feasibility. Since concepts are externally grounded, we simply do not have the power to change or replace concepts; this is the objection from externalism. The discontinuity objection is that replacing a concept does not fix the problem but merely changes the topic. The last objection against feasibility is the objection from fundamentality. The idea is that *truth* is just too basic of a concept to be replaced. Specifically, I consider the objection in the form of the argument that *truth* is inextricably entangled with *belief,* and that *belief* is fundamental to our psychology.

From there, we move on to objections against recommendability. The objection from centrality grows out of the fact that *truth* is a central concept, i.e. lots of other concepts are partly constituted by *truth*. If we operate under a principle of conservation or if conservation is something like a virtue, then the problem with replacing *truth* is that every concept partly constituted by *truth* needs to be replaced as well, and that is not at all conservative. Hence, replacing *truth* has a high theoretic cost. The next objection against recommendability is grounded in Mona Simion's account of conceptual normativity (Simion, 2018), i.e., the normativity that conceptual change, replacement, elimination, and conservation is governed by. According to Simion's account, conceptual normativity is grounded in preventing epistemic loss. If a conceptual change creates an epistemic loss, then we should not allow for that conceptual change. Replacing *truth* arguably leads to epistemic loss given the central role *truth* plays in epistemology. While this objection is less generalizable to concepts other than *truth*, it could

minimally be generalized to *belief* and *justification*. The objection from lack of justification is that *truth's* inconsistency (or any other problem *truth* might have) is either not a problem or that the problems are not worth the trouble of fixing.

In the third chapter, I present a theory of cognitive roles and tools, and I argue that once we understand *truth*'s cognitive role, *truth*'s replaceability becomes much more reasonable. I then argue that understanding how cognitive roles and tools allows us to address the objections presented in the second chapter.

There are a few things worth mentioning about methodology. First, this is self-consciously a work of analytic philosophy. That means I am writing for analytic philosophers, broadly construed, and that my usage of terms aims to align with the ways that they are used in analytic philosophy. Two, where I can, I try to carry out my arguments in a way that avoids committing to particular sides along the various fault lines in analytic philosophy. Ideally, I try to argue in a way that goes around debates rather than through them, i.e., make my arguments independent of any particular outcome of some debate. This strategy is evident in my discussion of replacement (supplantation) versus revision. That said, there are places I walk right in the middle of things and take sides.

1.0.0. Scharp's Argument for Replacing *Truth*

In *Replacing Truth*, Kevin Scharp introduces a novel approach to the liar paradox: replace *truth* with two concepts that will not give rise to a similar paradox. Here, I will give an overview of Scharp's argument for replacing *truth*. First, I will briefly review what the liar paradox is. From there, I present Scharp's argument that the liar paradox shows that *truth* is an inconsistent concept and that, unlike some other inconsistent concepts, *truth*'s inconsistency creates serious problems. Finally, I explain what Scharp's replacement concepts are and how

they avoid the paradox before saying something about replacement as a general strategy. The aim of this chapter is orient the reader by walking through a conceptual project, its mechanics, and its motivations.

1.1.0. The Liar and Inconsistent Concepts

Part of the charm of the liar is that it is surprisingly easy to understand its paradoxicality. There are many ways to formulate the liar, but I think the clearest is as follows:

L. *L* is false.

The problem is almost immediately obvious. If *L* is true, then "*L* is false" is true. But if "*L* is false" is true, then *L* is false. Consider *B:*

B. The boat is in the barn.

If *B* is true, then the boat is in the barn and so similarly for *L.* If *L* is false, then "*L* is false" is false. And since, under classical logic, when we deny that some sentence is false, we affirm that sentence's truth, *L* is true. If *L* is true, then *L* is false and if *L* is false, then *L* is true. Given the law of excluded middle, L must be true or false, so L is both true and false. Thus, we have arrived at a contradiction. Scharp takes this to show that *truth* is an inconsistent concept. Further down, I will put the whole argument in more formal terms, which will come in handy later.

What is an inconsistent concept then? According to Scharp, roughly "a concept is inconsistent iff its constitutive principles are inconsistent." (Scharp 2013, 36) Concepts have constitutive principles. These are the rules that govern the use and application of a particular concept. They, at least in part, make a particular concept the concept that it is. This is akin to the way games are at least partly constituted by their rules. While there are, perhaps, many ways in which a concept could be inconsistent, Scharp emphasizes inconsistency of application: when the

constitutive principles allow the concept to both apply and disapply to the same object. Consider 'groat':

1. 'groat' applies to *x* if x is a boat.
2. 'groat' disapplies to x if x is a green thing.[8]

'Groat' can be applied without inconsistency if it is applied to different objects. A red speedboat is a groat. A lime is not a groat. The problem is when we consider a boat that also happens to be green. Let us take John's green paddle boat. Since it is a boat, then it is a groat. It is green and hence is not a groat. John's paddle boat is a groat and it is not a groat: a contradiction. If we assume per *reductio* that green boats exist, then there is at least one object *x* that is both green and a boat. Since *x* is a boat, it is a groat. Since *x* is green, it is not a groat. Given the contradiction, the assumption must be rejected: there are no green boats. Scharp argues in (2013) and (2017) that many utilized concepts are inconsistent, including pejorative epithets, mass, (naive) infinitesimals, and set/membership.

To expand on the case of pejorative epithets, consider 'Okie.' We might think of the constitutive principles of 'Okie' as being:

1. 'Okie' applies if x is a person from Oklahoma
2. 'Okie' disapplies if x is not scum.

'Okie,' similar to 'groat,' could be applied without contradiction. There may very well be people from Oklahoma who are scummy. But there are definitely people from Oklahoma who are not scum. This latter group would be both Okies and not Okies: a contradiction.

[8] I structure groat on Scharp's rable.

What makes *truth* inconsistent? Again, for Scharp, an inconsistent concept arises out of its constitutive principles. The principles that Scharp (2013, 16) thinks make *truth* an inconsistent concept are the (T-In) and (T-Out) principles:

(T-In) If α, then ⟨α⟩ is true.
(T-Out) If ⟨α⟩ is true, then α.

The lowercase Greek letter 'α' is a sentential variable, and the angle brackets ('⟨' and '⟩') are used to form names (for our purposes quotations work as names). To give an example of (T-In), if Bob is bald, then "Bob is bald" is true, and for examples of (T-Out), if "Bob is bald" is true, then Bob is bald. If we think back to the liar, we can see these principles at work. Applying (T-In), if *L*, then "*L* is false" is true. By then applying (T-Out), if "L is false" is true, then *L* is false. *L* is true and *L* is false: a contradiction. We can now see how the application of both principles to the liar sentence (*L*) brings about a contradiction in a way, not unlike the application of *groat* to green boats. In both cases, we have a concept that can be applied without contradiction and only has contradictory consequences in exceptional cases. There is also a principle of substitution (2013, 16):

(Sub) If ⟨α⟩ = ⟨β⟩, then ⟨α⟩ is true if and only if ⟨β⟩ is true.

What this principle amounts to is that if the name of a sentence, a variable like L or a quotation, shares a referent sentence with another sentence, then one can substitute that name for the other. Now we give the more formal version of the liar:

1. L is true.	ASSUMPTION
a. 'L is false' is true.	SUBSTITUTION (From 1)
b. L is false.	T-OUT (From 1a)
2. If L is true, then L is false.	→ Intro (From 1 to 1b)
3. L is false.	ASSUMPTION
a. 'L is false' is true.	T-In (From 3)
b. L is true.	SUBSTITUTION (From 3a)
4. If L is false, then L is true.	→ Intro (From 3 to 3b)
5. L is true if and only if L is false.	↔Intro (From 2 and 4)

1.2.0. Scharp's Proposal and the Case for Replacement

Scharp's replacement consists of two concepts: *ascending truth* and *descending truth*. They approximate (T-In) and (T-Out), respectively. The basic idea is that we replace *truth* with two concepts, and by doing so one can avoid the alethic paradoxes as each of their respective constitutive principles cannot be used to bring about the paradox as the principle needed to bring about the paradox is part of the other concept. To clarify, we will go over how *descending truth* and *ascending truth* work.

With respect to the names of the concepts, it is 'descending' because one uses the corresponding principle, (D-Out), to move from a sentence about a sentence to the object sentence where metasentence is above and the object sentence is below. The latter is named 'ascending' because we move from what will be an object sentence to the metasentence using the corresponding principle, (A-In). This is ultimately used to build out a new semantics for "truth" that uses *ascending* and *descending truth* as well as the notion of safety that keeps our surface language the same while avoiding the liar.

Having summarized the project, we can now explain the mechanics of *ascending* and *descending truth.* The core principles of each are as follows (2013, 154):

(D-Out) $D(\langle\alpha\rangle) \rightarrow \alpha$
(A-In) $\alpha \rightarrow A(\langle\alpha\rangle)$

Where D is descending true, and A is ascending true. The relationship between the two operators (D and A) is similar to necessity and possibility in classical modal logic. If α is necessary (□α), then α is possible (◇α). α can be possible while not being necessary (~□α ∧ ◇α). Similarly, if α is descending true, then it is ascending true. Yet, α might be ascending true without being descending true. It is in this latter circumstance where α would be considered 'unsafe.' For Scharp, α is safe when α is either descending true or not ascending true (2013, 153).

(Safety) $S(\langle\alpha\rangle) \leftrightarrow (D(\langle\alpha\rangle) \vee \sim A(\langle\alpha\rangle))$

As it turns out, the sentences that lead to alethic paradoxes turn out to be unsafe. So now I can show how *ascending truth* and *descending truth* avoid those paradoxes.

(*P*) *P* is not descending true.
(*R*) *R* is not ascending true.

We can try to use liar reasoning to garner a contradiction. Let us assume *P* is descending true (the formal representation is in brackets):

1. *'P'* is descending true [$D(\langle P\rangle)$]

From there, we can substitute in 'P is not descending true' for its representative variable '*P'* and get:

2. '*P* is not descending true' is descending true [$D(\langle P\rangle)$][9]

Given the core constitutive principle of *descending truth* [(D-Out) $D(\langle\alpha\rangle) \rightarrow \alpha$], we can get the following:

3. *P* is not descending true.

[9] Scharp, for whatever reason does not seem formally distinguish in the syntax between a pure variable as a name and the quotation as name.

Since there is nothing like (T-In) for descending truth[10], we cannot infer from 3 that:

4. '*P*' is not descending true [~D(⟨*P*⟩)]

Nor can we infer with any number of steps the following:

3. *P* is not descending true.
.
.
.
n. It is not the case that *P* is not descending true [~*P*]

That said, it is arguably ambiguous whether there is a contradiction. Understanding things formally, there is not because P cannot be substituted for ~D(⟨*P*⟩), and the difference between ~D(⟨*P*⟩) and P prevents a contradiction from being expressed.

Further down, I will make a direct comparison between the traditional liar from above and the attempt to reproduce the liar using descending truth. There I will give a reading more favorable to a hypothetical opponent. But for now, we will try starting from the negation instead:

1. '*P*' is not descending true [~D(⟨*P*⟩)]

Without a (T-In) counterpart, there is not much we can do here. We cannot, for example infer:

2. ''*P*' is not descending true' is descending true [D{~D(⟨*P*⟩)}]

There is no way, to put it formally, to get:

~D(⟨*P*⟩) → (D(⟨*P*⟩) & ~D(⟨*P*⟩))

Alternatively, we can assume:

1. *P* is not ascending true [*P*]

[10] In Scharp (2013), he explicitly denies the following four principles:
(i) p is descending true → 'p is descending true' is descending true
(ii) 'p is ascending true' is ascending true' → p is ascending true
(iii) if p is a theorem of ADT, then p is descending true
(iv) if p is ascending true, then p is a theorem of ADT

Or even:

1. It is not the case that *P*

Again, without a counterpart to (T-In), we cannot infer anything that will get us the paradox. This is the key feature of Scharp's replacement. Earlier, I suggested there was an ambiguity due to the formalism. Someone could object that the only reason that a contradiction can be prevented is by hiding behind the formalism. That turns out to be only half 'true.'

As I mentioned earlier, I will make a direct comparison between a version of the traditional liar and an attempt at liar reasoning with descending truth without the supposedly protective formalism:

1) L is true.	ASSUMPTION	1) *P* is descending true.
a) 'L is false' is true.	SUBSTITUTION	a) 'P is not descending true' is descending true.
b) L is false (from 2)	T-OUT/D-OUT	b) P is not descending true.
2) If L is true, then L is false.	→ Intro	2) If P is descending true, then P is not descending true.
3) L is false.	ASSUMPTION	3) P is not descending true.
a) 'L is false' is true.	T-In/???	
b) L is true.	SUBSTITUTION	
4) If L is false, then L is true.	→ Intro	
5) L is true if and only if L is false.	↔ Intro	

Here, we do not distinguish between 'P' and '~D(⟨*P*⟩).' This allows us to draw a contradiction from assuming that P is descending true, but we can make no moves beyond that. Ultimately, the most we could show is that P is not descending true since we can draw a contradiction from assuming that it is true, and that P is ascending true which makes it an exemplar of an unsafe sentence (a sentence that is not descending true but is ascending true).

Let us turn our attention to R. Under Scharp's formal language, we can assume R, apply (A-In), then substitute and get:

1. R is not ascending true [R]
2. 'R is not ascending true' is ascending true [A(⟨*R*⟩)]
3. '*R*' is ascending true [A(⟨*R*⟩)]

But we have no way of dropping the ascending operator or negating 'A(⟨*R*⟩).' Formally, there is no set of inferences to get from '*R*' to '~A(⟨*R*⟩)], if we start with the assumption that:

1. 'R' is not ascending true [~A(⟨*R*⟩)]

We cannot even use substitution since Scharp's system requires that the sentence be true. Even we could substitute and get:

2. 'R is not ascending true' is not ascending true [~A(⟨*R*⟩)]

There is nothing we can do here without a counterpart to (T-Out) for ascending truth. Let us drop the formal mechanisms:

1) L is false. a) 'L is false' is true. b) L is true. 2) If L is false, then L is true. 3) L is true. a) 'L is false' is true. b) L is false 4) If L is true, then L is false. 5) L is true if and only if L is false.	ASSUMPTION T-In/A-In SUBSTITUTION→ Intro ASSUMPTION SUBSTITUTION T-OUT/??? → Intro ↔Intro	1) *R* is not ascending true. a) 'R is not ascending true' is ascending true. b) R is ascending true. 2) If R is not ascending true, then R is ascending true. 3) R is ascending true. a) 'R is not ascending true' is ascending true.

This is the converse of the *descending truth*. We can derive a contradiction from the negative assumption since we have a counterpart to (T-In) but missing the counterpart to (T-Out), we again are unable to pull out a contradiction and hence are unable to generate the paradox.

With both *descending* and *ascending truth*, regardless of which assumption we start with or whether we use metalinguistic formalisms, we do not generate the paradox. As suggested above, this is because descending truth has no counterpart to (T-In), and ascending truth has no counterpart to (T-Out). As such, they cannot be used to infer 'P ↔ ~P' nor 'R ↔ ~R.' Similarly, the paradoxes cannot be generated at the level of metalinguistic sentence, that is sentences about

sentences. The 'D' and 'A' operators are used to express metalinguistic sentences, but by assigning the (T-Out) and (T-In) counterparts to each operator respectively the paradox cannot be generated i.e. 'D(⟨P⟩) ↔ ~D(⟨P⟩)' nor 'A(⟨*R*⟩) ↔ ~A(⟨*R*⟩).'

An important thing to remember is that Scharp proposes only replacing *truth* where *descending truth* and *ascending truth* come apart, and they only really come apart where *truth* would result in a paradox i.e. unsafe sentences. As I mentioned in the introduction, Scharp thinks that part of his project uses *ascending* and *descending truth* as tools to explain the semantics of *truth* itself, using the replacements to explain the failures of the replaced. Going too far into this would lead this thesis astray, but it is worth keeping in mind both in respect to the scope of the project and the possible perk of adopting the replacement.

We might wonder whether *truth* needs replacing even if it is an inconsistent concept. Scharp claims that an inconsistent concept only needs to be replaced if that concept's inconsistency gets in the way of doing the job of that concept. In this case, Scharp claims that *truth*'s inconsistency undermines its explanatory role in truth-conditional semantics.

What is truth-conditional semantics, and what does it have to do with anything? A truth-conditional semantic theory will explain the meaning of sentences by means of their truth conditions: the meaning of a sentence α is nothing more than what conditions would have to be fulfilled for α to be true. As Scharp explains[11], truth-conditional semantics is a dominant form of semantic theory in linguistics and has lots of explanatory power. Scharp argues that even if it

[11] "The acceptability of truth-conditional semantics comes from linguistics, where it is firmly entrenched and has many explanatory and predictive successes, and the modest attitude toward the relation between philosophy and the sciences." Scharp (2013, 125)

turns out to be wrong, its successes require that the better theory explain why it worked as well as it did and that means explaining *truth* and the alethic paradoxes.[12]

Why exactly is the inconsistency of *truth* a problem for truth-conditional semantics? Here is a passage from Scharp on the matter:

> Any 'off the shelf' truth-conditional semantic theory for the truth predicate is going to be inconsistent. It will imply that a liar sentence is in the extension of the truth predicate iff it is not. So not only is the concept of truth inconsistent, its inconsistency impedes its utility.

I read Scharp here as saying that if *truth* is inconsistent, then the truth conditions of liar sentences are both met and not met: a contradiction. One way to think about Scharp's replacement is to think about it being custom-made for truth-conditional semantics. I take his proposal to address at least two theoretical problems with one stroke.

The problem with many of the approaches to the liar is that they must eventually restrict the truth predicate. If not that, they are forced restrict other parts of language that seem innocent. Even approaches that manage to avoid restriction initially often led to revenge paradoxes. Revenge paradoxes are liar-like paradoxes that can be built out of the resources that a theory uses to defeat or prevent the liar. A revenge paradox often leads to the need to restrict the language such that the revenge paradox does not come about. I show the basic strategy for building revenge paradoxes below.

Let us consider the classic strategy of restricting self-reference: we can say that self-referring sentences are meaningless. The first problem is that the following sentence is both self-referring and obviously meaningful:

(TNR) The font of this sentence is Times New Roman.

[12] It is not clear in Scharp's works on the subject why a replacement for truth would necessarily *have to* solve alethic paradoxes as opposed to providing some other meaningful improvement.

Sentences like this leave such approaches in a bind. (TNR) seems to be perfectly meaningful, i.e. it has semantic content. This might tempt one to ban sentences containing "true" and "false" from self-referencing but that is just arbitrarily picking out "true" and "false." Moreover, it seems that self-reference is not the problem but rather "true" and "false" since other predicates do not require the same restriction. From there, one might bite the bullet and just banish all self-referring sentences to meaninglessness.

Banishing all self-referential sentences to meaninglessness throws the baby out with the bathwater. There are plenty of self-referential sentences explicable in truth-conditional semantics that get restricted just to deal with the paradoxes. Let us say the theory is revised where “true” and “false” can only apply to sentences that can be evaluated externally i.e. where the truth value is not completely contained within the sentence. 'Snow is white' can be externally evaluated because one can look and see if snow is white.

(TNR) can be externally evaluated because its claim about itself is not true or false as a result of merely making the claim. You can evaluate the sentence by seeing whether it is in Times New Roman font or not. 'This sentence is true' is only true because it says of itself that it is true. We could call such sentences meaningless.

Such a restriction might help make for an attractive theory, but it runs into an immediate revenge paradox. Consider:

(D) All penguins are at least 10 feet tall, or D is false.

It can be externally evaluated. We can check if there are any penguins less than 10 feet. There are. Since the first disjunct is false, either the second disjunct is true or D is false. If the second disjunct is true, then D is false, but if D is false, then the second disjunct is true, which would make D true, leading us to evaluate both disjuncts.

We will again evaluate the first disjunct as false which would mean that the second disjunct is true since D is true, but affirming the second disjunct means D is false, and round and round we go. Assuming D is false just jumps to the second step of assuming that the second disjunct is true. D is not automatically false, because the first disjunct, if true, would make the sentence true. It is specifically because the first disjunct is false that it leads to the paradox. Consider:

(TNR+) TNR+ is written in Times New Roman, or it is false.

The first disjunct is true and externally evaluable, which makes the sentence meaningful. Of course, more could be said here, but I just wanted to give an example of how a revenge paradox could work.

Many revenge paradoxes take the form of '(X) X is either false or α' where α is what we might call the failure state of propositions i.e. propositions that fail to be true or false (gappy, meaningless, indeterminate, unstable, etc.). These failure states are states that many attempted solutions to the liar paradox assign to such propositions in an attempt to block the paradox.

From there, one proceeds to show that assuming X is true, 'X is either false or α' is true by substitution. From there, by T-Out, one can infer that X is either false or α. Hence, X is true if and only if X is either false or α. If we assume X is either false or α, then we can infer by T-In that 'X is either false or α' is true. By substitution, X is true which implies that if X is either false or α, then X is true. Since we have two conditionals that are the converse of one another, we can infer that X is true if and only if X is either false or α.

That means claiming X as false or as falling into whatever the failure state is means that X is true, and vice versa. This allows a liar-like paradox to reemerge[13]. Scharp's solution avoids

[13] This example is a generalized version of one that Scharp gives using the example of Kripke's Strong Kleen minimal fixed point theory (2013, 84-85).

this situation by splitting up T-In and T-Out into two different concepts which prevent the paradox and revenge paradoxes from emerging. One can see how this works with *ascending truth* in the following comparison:

L. L is false or meaningless.

R. R is not ascending true or meaningless.

1) L is false or meaningless.	ASSUMPTION	1) *R* is not ascending true or meaningless.
a) 'L is false or meaningless' is true.	T-In/A-In	a) 'R is not ascending true or meaningless' is ascending true.
b) L is true.	SUBSTITUTION	b) R is ascending true.
2) If L is false or meaningless, then L is true.	→ Intro	2) If R is not ascending true or meaningless, then R is ascending true.
3) L is true.	ASSUMPTION	3) R is ascending true.
a) 'L is false or meaningless' is true.	SUBSTITUTION	a) 'R is not ascending true or meaningless' is ascending true.
b) L is false or meaningless.	T-OUT/???	
4) If L is true, then L is false or meaningless.	→ Intro	
5) L is true if and only if L is false or meaningless.	↔ Intro	

As one can see, it works at preventing revenge paradoxes as well as the liar by blocking the inference one must make to generate the paradox. There is no equivalent to T-Out for ascending truth. Both descending truth and ascending truth are going to resolve revenge paradoxes the same way as they do on the liar.

This allows Scharp to have a complete truth-conditional semantics by creating a roughly truth-conditional semantics. That is an ascending-and-descending-truth-conditional semantics that works almost exactly like truth-conditional semantics except there is no temptation to restrict because neither the liar nor its respective revenge paradoxes are produced, and so the meaning of every sentence can be spelled out in terms of its ascending/descending truth conditions, even liar and liar-like sentences.

1.3.0. Replacing Truth as General Strategy

As I have already mentioned, for this thesis, I am primarily concerned with defending the feasibility and recommendability of replacing truth in general. Here Eklund states what I think of as intuitive starting points for a general argument:

> What might be an example of a truth-like concept that could serve to replace the ordinary concept of truth? If we look at the liar literature, we can find some concrete suggestions. To take a historically important example, Tarski's (1935) hierarchy of truth predicates was meant to serve as a replacement for the ordinary concept of truth, which Tarski deemed to be inconsistent. Moreover, in the contemporary literature, one finds a whole variety of theories – versions of Kripke's theory, various revision theories, various paraconsistent theories, etc. All of these theories describe possible concepts of truth, even if at most one of them can successfully describe the ordinary concept. Whichever ones of these theories fail to describe the ordinary concept describe possible replacement concepts. (2014, 294-295)[14]

I read Eklund as saying that, generally, solutions for the liar involve conceptions of *truth*, or *truth*-like concepts, different from the ordinary or naive concept of truth. If any concept proposed counts as being a replacement by being different than the ordinary or naive concept, then most philosophical projects are going to count as replacement projects since any philosophical work done on a concept is going to introduce changes. It will produce a new concept i.e. a replacement. Replacement, then, is nigh unavoidable in philosophy, and perhaps in intellectual work more generally, and that gives us a prima facie plausibility for replacement projects. I expand on this argument in section 1.3.2.

As I outlined in the introduction, I lay out the various objections in Chapter Two and provide a plausible framework in Chapter Three which avoids those objections, but must head off some distinction mongering that would make more work for us later. As mentioned previously, I have found that a somewhat common intuition is that revisions of concepts is both

[14] He references Tarski (1935), but that work will be listed under (Tarski 1953) as I used the one that was part of a later compilation.

more plausible and more acceptable than replacements. I imagine many worry that a lot turns on whether there is a modification of the existing concept versus a brand-new concept being put in the place of the old concept. To avoid conceding that revision is not replacement, I distinguish revision from supplantation, where revision suggests strong continuity and supplantation a strong break. I argue that issues surrounding conceptual identity are such that no useful distinction can be made, and that even if one could, there would be no decisive normative difference. I end the section and the chapter by arguing, as suggested above, that replacement is pervasive, and that should give replacement as a general strategy a prima facie plausibility.

1.3.1. Revision Versus Supplantation

In this section, I aim to accomplish a few things. The section comes in five parts. First, I lay out a plausible account of what a revision-supplantation distinction could be. Second, I argue that such a distinction based on the sort of endurantist or perdurantist models that apply to things like personal identity are unworkable. Third, I argue that a distinction based on a concept-conception distinction where conceptions are additive does not have much better prospects. Fourth, I argue that any normative preference for revision comes down to an appeal to conservation, and what value or normative advantage being more conservative brings can be overridden. Finally, having cast doubt on the viability of the distinction and normative decisiveness of conservation, I argue we should treat revisions as replacements in the sense characterized in the introduction.

One of the issues that Patrick Greenough (2017, 9) brings to bear on Scharp's general position is that it depends on a particular metaphysics of concepts, where concepts are identified by their constitutive principles in a relatively strict way. Greenough calls this the Concept Identity Principle:

(CIP): Concepts are individuated by their constitutive principles.

According to Greenough, the CIP fails to be true if an endurantist or pedurantist view of concepts is true. If concepts endure, they exist wholly in the present but are allowed to change while remaining the same thing. If concepts perdure, they exist as the sum of their temporal parts. They both allow a kind of change of concepts. That is concepts can change while remaining the same concepts. Both views would allow concepts to be merely altered without being supplanted by a new and distinct concept.

We should begin by distinguishing between revision and a narrow sense of replacement that I call supplantation. There is an important sense where both alterations of concepts and genuinely new concepts would come to replace older concepts. If an alteration of a concept, a revision, becomes the dominant version of a concept, it does replace the older version. In other words, newer conceptions replace older conceptions. I will expand on this below. This is where we want to distinguish replacement from supplantation. A supplantation would be unlike a revision because it would be considered a genuinely new concept based on some kind of criteria. It might initially seem that a relatively strong distinction could be made between revision and supplantation. However, it turns out that there are several difficulties in trying to distinguish between revisions and supplantations because so much depends on the identity conditions of concepts.

What should be made apparent is that when we say that a revised concept is the same concept as the concept revised (i.e. the original concept), the sense of sameness employed here is somewhat loose. The revised concept is the same concept in the same way I am the same person I was in 2005. There is, perhaps, a sense in which I am the same as I was in 2005 but there are also rather significant differences (my back hurts way more often for example). It is these kinds

of intuitions that endurantism and perdurantism about the identity of objects and persons are supposed to capture. When two things are the same in this sense, it is simply to say that they share something special that makes them count as the same or there is enough continuity to count as the same.

The sameness here does not correspond to the identity relation used in first-order logic (i.e. if $x = y$, then for everything predicable of x is predicable of y also). Part of what makes a revision a revision is that it is somehow different from the thing revised. There has been some kind of change made. This implies that the original concept and the revised concept are not the same on a strict notion of identity as they fail to share all the same properties. One issue is that any theory of identity conditions for concepts is going to be contentious, and what counts as a revision and what counts as a supplantation are going to depend on the accepted theory. I will in this section explain why pedurantism and endurantism about concepts are unlikely candidates for conceptual identity, but also why on any theory of conceptual identity revision and supplantion will not be decisive.

It seems unlikely that any kind of decisive difference is going to result. Whether any kind of purposive conceptual change is both possible and permissible, whether in terms of revision or replacement, is a separate question from the question of the distinction between revision and replacement. If purposive conceptual change is possible, then presumably we will know of at least some cases where purposive change is both able to be done and should be done and it will likely be unsettled whether the new concept or concepts are revisions or replacements. Yet since we know that they are to be implemented then it seems unlikely that their to-be-implemented-ness depends heavily on whether they are revisions or replacements.

Any theory that models the identity of concepts on something like personal identity or the identity of physical objects will have to address a genuine issue. Namely, the original concept does not go away. When I change, newer versions of me completely displace the old versions. There remains a single entity. This is not so with concepts. One can arguably revise them, but concept X.1 and concept X.2 will both be intelligible independent of one another. More than that, a concept can be revised multiple times, in something like a branching structure. I might revise Concept X, call it X.J, and some other person might revise X, call it X.O., thus creating two separate branches off the original concept X. That is three concepts which in some sense are one, but in another sense 3. These concepts, X.J and X.O might themselves be revised and then we might have something like six concepts at play: X, X.J, X.J.2, X.J.3, X.O, and X.O.2. This would be X branching off as .J and .O, and then .J and .O branching off into their own .2 and .3, respectively. We might be tempted to invoke something like concept pluralism here, but the issue is that the perdurantist and endurantist will insist that there is specifically one concept, where what I have done is named either different past points of the concept or different temporal parts.

If a newly revised concept is still the same concept, and revisions of a revision continue the identity, then this means we have six versions or variations of one concept. If I did something similar to myself, where modified copies of me were created and each of the me's were coexisting, I very much doubt that the different versions of myself would be counted as being the same thing. Thus, the best interpretation of the names I gave (X, X.J, X.J.2...) is that they name different branches of the self-same object. This suggests that personal/object identity is not a good model for conceptual identity: if we use personal identity as our model and we would not count multiple synchronic versions of a person as identical then we would not count multiple versions of a concept as identical. Plus, these metaphysical models make concepts needlessly

complex and metaphysically heavy. They create more questions than they answer[15] and they have a real "how many angels can dance the head of a pin" flavor to them. The CIP in comparison is much more straightforward where we do not need to, for example, answer questions about how different versions of supposedly self-same concept related to each other. We just have a new concept any time we have a difference in constitutive principles.

If we want a theory of enduring and perduring concepts, a different model is in order. Perhaps, we should not think of revisions of concepts in terms of identity at all. If one concept can have endless variations and still be *one* concept, we might wonder whether it is one in any meaningful sense rather than just being one by decree. This is at least one reason we might be skeptical of some metaphysical difference between revisions and replacements.

We could employ one commonly used distinction, the concept-conception distinction. While the nature of the distinction is as disputed as anything else in philosophy, conceptions are often thought to be a further specification of some concepts. One way that the distinction could be filled out that gels with this work is that a concept has core constitutive principles that conceptions build on with further principles or the modification or removal of non-core principles. The distinction laid out this way does arguably map onto the supplantation-revision to, at least, a considerable extent: supplantations are new concepts insofar as the core principles are changed and revisions insofar as the change to the concept is additive or non-core principle are removed. This could give us what seems to be a clean answer to how the X family of

[15] To list: exactly what sort of change constitutes a branch, a change in constuitive principles, a change the associated word or term, a new connotation of the associated word or term? If some far branch somehow comes to be constittuted the same as the trunk, does the branch merge with the trunk? Is merging in general possible? Can concepts or their branches die or end? How should with think of the relations between branches with each other and their trunks, especially when it comes to their usage? Are branches necessarily distinct or can they be fuzzy?

concepts are both one and many, one concept and many conceptions. We might wonder how much the distinction helps us, however.

One issue is the matter of determining which principles are core and which are non-core. Are (T-In) and (T-Out) core or peripheral? What is it that makes a principle core? Finding a way to cleanly differentiate between core and peripheral that is not arbitrary and that would track some normative difference is a tall task. The theory would to be able to tell us what is and is not a core principle, but also explain how the difference between core and peripheral is such that revisions being conceptions makes them favored over supplantations being new concepts. More than that, that favored status must be decisive: The revised conception is always or almost always favored over the supplanting concept, if supplantions are allowed at all. I cannot rule out such a theory, but the mere possibility does not need to be answered for.

Another issue is that it is not obvious whether the distinction is absolute or relative. The distinction might be more like genus-species in the logical sense in that it describes a relation. To use my previous example X is a concept relative to its conception X.J but X.J is a concept relative to its conception X.J.1. The concept-conception distinction could be absolute. For example, maybe concepts are always employed through the medium of conceptions, that concepts are at least one step removed from us. We work backward from our conceptions to determine which concepts we are familiar with fall under. One can see that in the former case, it would be difficult to stake any general normative difference between concepts and conceptions if any concept or conception might be a concept or conception in a different context. The latter case would seem to allow for distinct norms given they are distinct things. The latter case seems less plausible if only because it seems that an added principle could become core in the future,

that a conception could become a concept under which new conceptions could fall. Of course, perhaps my humble proposal for the distinction is a failure for those reasons.

Here is a chance to dive deeper into the normative and evaluative reasons for changing concepts while we consider whether there would a decisive normative difference even if we had a genuine distinction. Maybe the issue is that, by their very nature, revisions are the more conservative choice with respect to supplanting concepts and we should seek to conserve whenever possible. There are many reasons to doubt such a principle, but if we are to grant it, it does not mean that revisions are always to be preferred to supplanting concepts. Even if a revision's status as more conservative were to give the revised conception more value or some kind of normative advantage, this greater value or normative advantage can surely be overcome by other considerations if strong enough.

It might be that although supplanting a concept is permissible, it very much is the exception. Only under very special circumstances should we supplant rather than revise. Is there any reason to think this is the case? Arguably, when we think about conservativeness, we are thinking of it as a stand-in for a collection of other values or considerations anyhow. Generally, there are at least two things that conservativeness stands in for, lesser resource-intensiveness and less risk. There are theoretical versions of these. Inquiry, even philosophical inquiry, takes at least time and effort. Consider some theory that is, on the whole, epistemically successful but has some problem areas. We could replace the whole theory, or we could make small modifications to the theory to solve the problems. The latter is generally preferable because small modifications take less time and energy than working out a whole new theory. Furthermore, we tend to think that what we are familiar with (an old theory perhaps) is less risky than what we are not familiar with. A new theory might have all sorts of epistemically undesirable implications.

This might be why philosophical theories thought dead arise anew with contemporary reworking. Resuscitating an old theory is generally easier than producing a whole new theory, and since it has already been put through its paces, there are fewer surprises.

There are situations where the problems are just too big or too deep, and the theory just needs to be scrapped or modified so much it becomes unrecognizable. The properties of being less resource-intensive or less risky do not automatically make beliefs, theories, or concepts the best among contenders. In some cases, a newer concept being more like the target concept is a mark against the newer concept. If the problem lies in the targeted concept, that might be a reason to try and make the new concept, whether a revision or a supplantation, sufficiently different from the targeted concept. While lesser resource-intensiveness and less risk often give revisions a normative edge, these are reasons to think that cases where supplantation is the better option are not especially rare. Again, this is assuming that a neat distinction could be made between the two.

The distinctions we have explored between revision and supplantion that are maybe substantive enough to ground some decisive normative are seemingly unworkable, and that conservation alone is not decisive. Given these, we ought to consider that no distinction is necessary. Even if it could be shown that there was some way to distinguish between revisions and supplantations, it would still be the case that lots of variations or versions of a single concept are possible. We would still have to decide which revision (or branch or conception) we would want to use, which version should usurp the original concept. This would not be decided by the fact that one is a revision, and the others are supplantations since there would be other revisions to choose from. There are a lot of diverse ways in which one revision might be better than another: parsimony, more predictive power, more explanatory power, solving more conceptual

defects and theoretical problems (such as paradoxes), more appropriate scope, greater theoretical unity, good consequences, and so on. The question would be why we could not simply use these same means of comparison for supplantations. If none of the proposed distinctions are workable or are substantive enough to be normatively decisive, then there is no reason to think we could not measure supplantations by the same yardstick. We do not even really need to consider revision versus supplantation because we can simply consider how different the proposal is and whether its advantages outweigh any weight we might put on conservation (if any).

Ultimately, any change is going to count as replacement here. The point of both revision and supplantion is to make some kind of improvement, and for there to be an improvement there has to be a some meaningful difference from the concept to be improved on. Whether a new conception takes over from an old conception, a revision modifies the concept, or an old concept is supplanted by new concept, something is being replaced. A conception, a revised concept, a supplantion is replacing an older conception, an unrevised concept, an old concept respectively.

To sum up, since a neat distinction seems unworkable, would not make a normative difference if it were workable, and we replace regardless of what is doing the replacing, we ought not bother with a distinction between revision and replacement. Revision, if it is anything, is replacement. What we are ultimately doing is weighing a number of different values, problems, and other normative factors when we are considering replacement and what does the replacing. If replacement is something we can and should be engaging in, this is roughly all we need to consider.

1.3.2. The Pervasiveness of Replacement

We have a picture of what a conceptual engineering project looks like and what motivates it. We can imagine, I hope, the shape of what other such projects might take. We have some

sense of what might be at stake. Here I want to give the pervasiveness of replacement the air of plausibility if it does not have it already. Part of the work has already been done by showing that broadly that there is only one activity that concerns us, not two: replacement, not replacement AND revision. First, I want to explain what happens in an introductory philosophy class in terms of replacement. If it explains well, then mere plausibility should come for free. I will start with a formal argument that I can refer to and explain as I move through the rest:

1. Regardless of what continuity a concept or conception might have with what has come before, it only makes sense to make a proposal if that concept behaves, so to speak, in a way that addresses whatever the issue is while having reason to believe the older concept could not.
2. However, it is that difference in conceptual behavior that any substantive difference is constituted, the adopting of the newer concept or conception counts as replacement in the minimal sense.
3. Proposals for alternatives with respect to received concepts are common enough in philosophy to be encountered in most, if not nearly all, branches and subbranches.

∴ Given 1, 2, and 3, proposals for replacement are pervasive.

Regardless of whether we distinguish between revisions and supplantations, or concepts and conceptions, we nonetheless replace. In academia at least, when a sentence is true, it is not false. There is one kind of being true, not two or more. Saying a sentence or a set of sentences is true is to endorse those sentences. If something is not true, then it is false (contexts involving vagueness might be an exception here). *True* applies to sentences (or propositions) and no other kinds of things. A sentence being true is not subject-relative. This supports premise 1.

At least until you ask someone on the street. The notion of *true* I have described might be described as *scholarly true*. It is, more or less, the use of "true" that gets taught in introductory philosophy class. It arguably at least approximates how "true" is by scholars across disciplines (science, math, humanities, etcetera). Introductory philosophy students often resist *scholarly true* in a number of ways: what is "true" me for might not be "true" for you, sentences can be kind of "true," the "true" of religion and the "true" of science is not the same, among others. There is

often both agreement and dissent with *scholarly true* among the competing notions of "true." This is all to say that insofar as there is a naive concept picked out by "true" then there is more than one. That, or *scholarly true* is not the naive notion.

The various versions of "true" that have been utilized by the various attempts to solve the liar will similarly disagree with *scholarly true*. Paraconsistent solutions allow a sentence to be both true and false. For Tarski, there is an ascending hierarchy of truth predicates i.e. there is always some further truth predicate that can be used to say of some "is $true_n$" that it "is $true_{n+1}$" where the latter is higher in the truth predicate hierarchy than the former. One also finds a regular menagerie of uses of "truth" and truth values: *determinate truth, definite truth, stable truth* as well as truth values such as *meaningless, unstable, undefined, sub-true*, and so on[16]. All these variants do not line up with *scholarly true* but also are more structured and specified than the lay concepts of named by "true."

We reasonably say that when we teach the way analytic philosophy (for example) uses "true," we are often teaching students to replace the way they use "true." That is the concept or conception of truth they have used gets replaced by *scholarly truth.* Sometimes it sticks and sometimes it does not. However, it does stick at least some of the time, and for good reasons. Even if not ultimately consistent, it does not bear incoherence so closely to the surface as a "true for me but not for thee" variation does. It helps create a common discourse for scholarly work, a common game board so to speak. These uses of "true" worked well enough for these students before adopting *scholarly true*, the same way *scholarly true* works for most philosophers despite that it might ultimately be inconsistent. We try and replace *scholarly true* for the same or similar

[16] You can find *infinite truth* (my name) in (Tarksi 1953, 194-195), *determinate truth* in (Field 2008, 153-155), *definite truth* in (McGee 1990, 150-151), *stable truth* in (Gupta and Belnap 1993, 194-203), *meaningless* in (Carnap 2003, 326), *unstable* in (Gupta and Belnap 1993, 173), *undefined* in (Kleene 1974, 332-333), sub-true (Hyde 1999, 733).

reasons we replace the lay variations. It is replacement, because however we draw our distinctions, there are crucial differences. They are significant differences that alter the context of their use, their functionality, the norms of their use, and so on. What is replaced by *scholarly true* is different from *scholarly true*, and whatever replaces *scholarly true* is different still whether that be a difference in constitutive principles or something else. This supports premise 2; the difference that makes a difference makes the new conceptions different enough to count as a replacement.

Philosophy can be seen as full of replacement. We can read the compatibilist attempts at a concept of free will to create a concept that is less metaphysically heavy, that aligns better with the best science of the time. Attempting to solve the Gettier problem can be read as an attempt to find a suitable replacement for *knowledge as justified true belief*. This view of philosophy arguably makes it more continuous with other forms of inquiry. Under this view, concept creation is analogous to hypothesis creation. We try on different concepts and test them against our intuitions or expectations, on their functionality, on their explanatory power, on their consequences, and so on. When things go well, there is a refining of the competing replacements due to the exchange of feedback. This supports premise 3. Insofar as this describes our practices, the simplest way to characterize them is using the notion of replacement. Differences that make a difference in our concepts or conceptions make our proposed concepts' or conceptions' replacements with respect to what was accepted before. At least, replacement is what happens when a proposed concept succeeds. This is my expansion of the conclusion (∴).

Here I can finally fully cash out my promise from 1.3.0. If the above is a robust argument for what Eklund suggested, then the normative implication is as follows. If replacement is pervasive in philosophy and philosophy is a legitimate discipline, then it follows that

replacement is generally legitimate where legitimate means something like epistemically worthwhile. If most of philosophy involves proposing, arguing for and against replacements, then it would seem philosophy's legitimacy hinges on replacement's legitimacy. My argument makes a case for the first half of the antecedent and if the reader has read this chapter and understood it, then they most likely accept the second half. One could deny the premise by arguing that philosophy is legitimate despite replacement. One would, of course, want arguments to show that replacement is illegitimate, either because we cannot replace or should not replace. In the next chapter, we will review some of those arguments.

The Objections

2.0.0. Against Replacement

Here, I repeat the Chapter Two introduction from 0.0.0. We start with three objections against feasibility. Since concepts are externally grounded, we simply do not have the power to change or replace concepts; this is the objection from externalism. The discontinuity objection is that replacing a concept does not fix the problem but merely changes the topic. The last objection against feasibility is the objection from fundamentality. The idea is that *truth* is just too basic of a concept to be replaced. Specifically, *truth* is inextricably entangled with *belief,* and that *belief* is fundamental to our psychology.

From there, we move on to objections against recommendability. The objection from centrality grows out of the fact that *truth* is a central concept, i.e., lots of other concepts are partly constituted by *truth*.[1] If we operate under a principle of conservation or if conservation is something like a virtue, then the problem with replacing *truth* is that every concept partly constituted by *truth* needs to be replaced as well, and that is not at all conservative. Hence, replacing *truth* has a high theoretical cost.

The next objection against recommendability is grounded in Mona Simion's account of conceptual normativity, i.e., the normativity that conceptual change, replacement, elimination, and conservation is governed by. According to Simion's account, conceptual normativity is grounded in preventing epistemic loss. If a conceptual change creates an epistemic loss, then we should not allow for that conceptual change. Replacing *truth* arguably leads to epistemic loss, given the central role *truth* plays in epistemology. While this objection is less generalizable to

[1] This objection can be found in Patrick Greenough (2017).

concepts other than *truth*, it could minimally be generalized to *belief* and *justification*. The objection from lack of justification is that *truth's* inconsistency (or any other problem *truth* might have) is either not a problem, or that the problems are not worth the trouble of fixing.

I do not argue for these objections as thoroughly as I might have. Each one could easily take a chapter of its own. My goal is to present the objections in a way that balances expediency and strength. My aim is to show that there are serious worries for the advocate of replacement. If I show they are objections worth answering by the advocate of replacement, then I have succeeded in my goal. In the corresponding section of each objection, I try to make the sort of arguments and appeals that I think a person making *that* objection would make. Finally, it should be kept in mind that I will answer these objections in chapter three, or at least minimize potential worries.

2.1.0. Arguments Against Feasibility

Before launching into the various objections to feasibility, I want to take a moment to say a bit more about feasibility. Again, this essay attempts to answer whether *truth* can be the target of a conceptual engineering project. As I mentioned in the introduction, there are both practical and theoretical flavors of feasibility. Here, I will be almost exclusively focused on theoretical feasibility. There are indeed fascinating questions surrounding practical feasibility. However, many of the questions surrounding practical feasibility are empirical in nature and hence beyond the scope of this thesis.

One way to frame theoretical feasibility is in terms of whether conceptual engineering is metaphysically possible. We can ask whether there is something about minds, concepts, inquiry, language, or something about the target concept in question, such that no replacement is possible. As presented in the introduction, a conceptual engineering project is infeasible if there

are strong in-principle reasons to believe that it cannot be done. We can break this condition down further.

The best way to understand in-principle reasons against some possibility is to distinguish them from contingent reasons against a possibility. A contingent reason against a possibility will be conditional or based on some contingent fact. Consider a human throwing a baseball at 115 mph. This speed will be outside the range of even the best pitchers. There are reasons based on the anatomy of the human arm and body to think that no unaltered human will ever pitch that fast. However, mechanical or biological modifications of humans might make such speeds possible or even common. Those anatomical reasons would then be contingent reasons, given that our anatomy is ultimately changeable.

In contrast, there are likely throwing speeds unreachable by anybody reasonably classified as human. There may even be speeds that surpass the limits inherent to the nature of throwing. At the very least, the laws of physics would provide in-principle reasons against the possibility of humans throwing anything faster than the speed of light. Someone might object that the speed of light is contingent, given that the laws of physics are contingent. I am inclined to agree, but we need a positive reason to think that a human could both live and throw baseballs in a universe where the speed of light was not a limit. Relatedly, someone might point out that I am speaking about laws of physics as opposed to laws of logic or metaphysics. I am primarily interested in whether purposive conceptual change can be metaphysically accomplished. First, nothing about the laws of physics as I am familiar with them, seems to rule out purposive conceptual change; it seems it can be physically accomplished. Two, whether it is in fact, physically accomplishable is an empirical matter. That said, I think of “in-principle reason” and "contingent reason” as being relative terms in that, insofar as a reason covers more contexts,

more time, and is less contingent, it is an in-principle reason. Insofar as it is more particular, covers less time, covers fewer contexts, it is a contingent reason. Given the usual hierarchy of possibility, metaphysical reasons against some possibility would be considered in-principle reasons compared to physical reasons against that same possibility and contingent relative to logical reasons against the same possibility.[2]. I will specify how each of the following objections challenges the feasibility of replacing *truth* and other basic concepts.

2.1.1. The Objection from Externalism

Again, one objection to replacing concepts is that if semantic externalism[3] is true, then we cannot engineer concepts, thus not replace them, because how language is used is not up to us. Semantic externalism holds that concepts and their usage are grounded, in part, in facts outside of our mental states[4]. Burgess and Plunkett layout semantic externalism quite nicely:

> The textbook externalist thinks that our social and natural environments serve as heavy anchors, so to speak, for the interpretation of our individual thought and talk. The internalist, by contrast, grants us a greater degree of conceptual autonomy. One salient upshot of this disagreement is that effecting conceptual change looks comparatively easy from an internalist perspective. We can revise, eliminate, or replace our concepts without worrying about what the experts are up to, or what happens to be coming out of our taps. From the externalist's point of view, however, conceptual revolution takes a village, or a long trip to Twin Earth. (2013, 1096)[5]

If concepts cannot be revised or replaced without controlling those anchors that determine concepts, and we cannot control or strongly influence those anchors, then we cannot conceptually engineer any concepts, let alone *truth*.

---

[2] I think this would hold in other forms of possibility deeper in the nesting if there are other such forms.
[3] Sometimes, it is also referred to as metasemantic externalism. As far as I can tell, they are used interchangeably.
[4] It seems like, depending on exactly how we understand mental states, one could think that even mental states could count as external if, for example, those states are not consciously accessible.
[5] This passage was brought to my attention via (Simion and Kelp 2019, 994).

Of course, the conceptual engineer could simply try and refute semantic externalism. While an option, it is no small task. Even if one could mount an effective attack on the position, it is unlikely that it would be some knock-down argument or in any other way final and conclusive.

Insofar as that is the case, the proponent of conceptual engineering should find a way to address the problem less dependent on the outcome of a semantic externalism vs internalism debate or at least find a way to allay the worries that the externalist has. The threat that semantic externalism poses to feasibility is pretty straightforward. If we have little control over the reference-changing facts, then successful purposive conceptual change would be more or less just a matter of luck, if possible at all.

One way of explaining the problem is by using Cappelen's analogy between language/concepts/meaning and the Marxist idea of base and superstructure:

> Think of the metasemantic base as the grounding facts for meaning and reference. Think of the metasemantic superstructure as consisting (at least in part) of our beliefs, hopes, preference, intentions, theories, and other attitudes about meaning and reference (what they are and what they ought to be). (Capellen 2018, 58-59)

The grounding facts for meaning and reference are the non-linguistic physical, biological, and social facts that determine or fix the semantic facts.

The relationship between the base and the superstructure is asymmetrical: the influence that the superstructure (beliefs, theories, etc.) has on the base (physical, biological, and social facts) is significantly less than the influence that the base has on the superstructure. It could be that the superstructure's influence on the base might be outright minuscule. For example, Capellen explains his view thusly:

> Presenting an argument for a particular conceptual change has no more effect on conceptual change than presenting an argument for how to lower crime in

> Baltimore has an effect on crime in Baltimore. Both have no effect whatsoever. (Capellen 2018, 60)

Capellen argues for two specific points: an epistemic point and a metaphysical point. His epistemic point is that, even though change in meaning is possible, we have no idea how to do it. It is a messy, complicated, and often inconsistent process. As Capellen points out, even if we grant that meaning is based on use, we have no "recipe" for getting meaning from use.

There is not even a theoretical formula such that we could alter the way people use words and get a particular meaning, let alone having some practical method to induce such a change. This limitation would apply equally to changing what concepts are called forth by the words we use. It may just be too complex and arbitrary for there to be anything like a recipe, at least for humans.

Worse yet, even if we knew such a recipe, we likely have no way of implementing such a recipe. The things that fix our language and concepts are just going to be out of our control. That is not to say that these things cannot change, but rather that **we** cannot, for the most part, **purposefully** change those things in any controlled way. Any changes that we might be able to induce may not have the intended effect. As Cappelen explains:

> ...it is an illusion to think that any individual or group has any significant degree of control of the reference-fixing facts. If we are not in control of the reference-fixing facts, then we're not in control of conceptual engineering because it requires us to change the reference-fixing facts. Even if we were perfectly coordinated as a group (something we are decidedly not), we would not give the group control because the actions and intentions of groups have at best a messy and unpredictable effect on our semantic values. We can of course try to influence other speakers and experts- but that will hardly ever amount to more than a drop in the ocean. (2018, 74)

Being clueless and out of control, we cannot expect to be able to perform any kind of conceptual engineering.

2.1.2. The Objection from Discontinuity

The discontinuity objection can be traced to Peter Strawson objecting to Carnap's conceptual explication, which can be understood as an early form of conceptual engineering. Strawson says:

> [I]t seems prima facie evident that to offer formal explanations of key terms of scientific theories to one who seeks philosophical illumination of essential concepts of non-scientific discourse, is to do something utterly irrelevant – is a sheer misunderstanding, like offering a text-book on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart. (1963, 504)

Even contemporary conceptual engineers have this worry, such as Sally Haslanger:

> Revisionary projects are in danger of providing answers to questions that were not being asked. Given the difficulty of determining what 'our' concept is, it isn't entirely clear when a project crosses over from being explicative to revisionary, or when it is no longer even revisionary but simply changes the subject. (2000, 34)

What exactly is the objection? The issue is that when we engineer some concept, especially when we replace some existing concept, we might instead be changing what it is that we are talking about. As a result, we might not be solving the problem that we set out to solve but instead be creating a new topic altogether.

A recent instance of this is when Lawrence Krauss, in a *Universe from Nothing*, defined "nothing" as an absence of physical objects (2012, 22-24). He was criticized for this definition because he assumed that empty space, in which particles pop in and out of existence, still counts as nothing. Here, we can see the dispute about a change of topic. Krauss defines "nothing" in a way that he thinks that other physicists would agree is nothing, as opposed to a more robust sense of nothing. If Krauss merely wanted to argue that the universe came from nothing in Krauss's sense, he may well have succeeded, but if Krauss's goal was to undermine contemporary cosmological arguments for God's existence, he most likely failed. Krauss failed because the sense of "nothing" that those arguments depend on is not the sense that Krauss is

using. He changed the topic to effectively argue that "the universe came from nothing," i.e., not from God.

The way this applies to *truth* specifically is that if we replace *truth* for, let us say, the purposes of solving the liar's paradox, we might reasonably ask whether we have actually solved the liar's paradox, given that the liar's paradox concerns *truth* and not *ascending* and *descending truth,* for example. If we are merely changing the topic by replacing the concept of *truth*, then we have not solved the problem. One might wonder whether changing the topic is so bad. The problem is not a change of topic per se, but like the Krauss example, if I make an argument and someone attempts to refute that argument by changing the topic, I can reasonably say that person has not refuted my argument. At best, if conceptual engineering always involves changing the topic, then the scope and potential of conceptual engineering at least become much, much smaller. At worst, there is no conceptual engineering: there is just changing the topic, and that leaves conceptual engineering theoretically infeasible.

2.1.3. The Objection from Fundamentality

Another objection to feasibility is that *truth* is just too basic of a concept to replace. Replacing *truth* is like trying to substitute something for sea bass in fried sea bass. There simply is no fried sea bass without the sea bass, and there is no replacement without *truth*. Here is Matti Eklund on the issue:

> One problem we run up against here has to do with the fact that the concept we are considering replacing is, in some way, a very *basic* concept. What we are considering is the possibility of some concept C of representing the world correctly such that C is distinct from the ordinary concept of truth and somehow better captures the idea of correct representation of the world than the concept of truth does. One may easily be skeptical of the project, wondering what independent grasp we have of correct representation such that correct representation can be held to come apart from truth. (2014, 299)

If we think of *truth* as correct representation, as Eklund suggests, it could be argued that if we can replace concepts at all, replacements have to help create more correct representations. Cognitive achievement is marked by *truth*, by correct representation, and *truth* serves as the only standard by which a cognitive engineering project can be judged, just as it is in all of our cognitive endeavors. While several arguments could emerge from this general worry, I will articulate an argument that falls out of a commitment to the psychological reality of belief and certain views on belief's relationship with *truth*.

The gist of the argument is as follows. If, as humans, beliefs are unavoidable, and beliefs are normatively or conceptually tied up with *truth*, then that commits us to *truth*-seeking. If that is the case, then replacement is not feasible, because we cannot help but be *truth*-seeking critters. There are two reminders before I explicate this little argument. First, this thesis is self-consciously trying to work in the paradigm of analytic philosophy. Second, I am trying to avoid commitments to particular views. What these mean together in this particular case is that I am trying to stay neutral, relative to analytic philosophy as to what the metaphysics of beliefs is. Furthermore, while in this section, I am endorsing what some will recognize as evidentialist views of belief.

To expand on the quick summation, I will first lay down more precisely what our premises are:

(IT) Beliefs are an ineliminable aspect of human psychology.

(ET) Belief, one way or another, is fundamentally entangled with *truth.*

While (IT) and (ET) strongly worded, I suspect that they are at least widely plausible to at least analytic philosophers, with the exception of the eliminative materialists and perhaps some

adjacent positions[6]. So, it might be best here to look at what should be a relatively neutral definition of belief per the SEP:

> Contemporary Anglophone philosophers of mind generally use the term "belief" to refer to the attitude we have, roughly, whenever we take something to be the case or regard it as true. (Schwitzgebel 2023)

Even the introduction to the article about belief in the *Stanford Encyclopedia of Philosophy,* in its relative neutrality, ties *belief* up with *truth*.

The argument is simple, then. If human beings are such, they cannot do otherwise than have beliefs, and beliefs are inherently entangled with *truth*, then *truth* cannot be replaced, because *belief* cannot be revised or supplanted. We cannot get rid of *truth,* because we cannot get rid of *belief*. It is just how were built. Of course, this argument relies on evidentialist-adjacent intuitions and arguments. For the purposes of this paper, I don't need to convince anyone of (IT) and (ET). I merely need to acknowledge and eventually address the kind of argument the philosopher with evidentialist-adjacent intuitions might make.

2.2.0 Arguments Against Recommendability

As noted in the introduction, this essay is not merely concerned with whether we, in a strong sense, **can** replace *truth,* but also whether it is possible that we *should*. The present work seeks to show that when a presently used concept is found to be defective or deficient, we ought to strongly consider replacing the present concept with a less defective or better concept. Again, conceptual engineering, even basic concepts, is not only feasible but recommendable. Like the section on feasibility, I want to break down what I mean by recommendability.

[6] Eliminative materialists that the entities of folk psychology (beliefs, thoughts, etc) should be eliminated from our ontology and the language of science. (Ramsey 2022)

Recommendability, as I am using it, is prima facie deliberative considerability. There are two major elements here to discuss: *prima facie*-ness and deliberative considerability. I will begin with the latter. As suggested in the introduction, deliberative considerability is the property of a possible course of action, where if a course of action x is deliberatively considerable, it should be evaluated in relation to other possible courses of action either by an individual or a group.

Contrast this with a course of action that is infeasible, fails to relate to the given context properly, or is too risky and too costly given the context. We should again consider someone with a severe heart problem. Transferring the person's mind to a new body is, for the foreseeable future, functionally impossible and hence, infeasible. Throwing a party fails to bear relevance. Under the conditions where artificial hearts do work but are not always reliable and have lots of complications related to them, the course of action has little deliberative considerability (although some) due to the risk and costs involved. It is only to be considered when more considerable courses of action are mostly ruled out or after some failed attempts at other solutions. As these risks and costs diminish, replacing bad hearts using artificial hearts gains deliberative considerability up until the point that it should be one of the first courses of action considered.

Care should be taken not to conflate the descriptive and the prescriptive here. Here, we can distinguish the subjectively apparent courses of action from the deliberatively considerable ones. These will often overlap, but mental and cultural inertia, for example, might bring about cases where newly possible courses of action that are low-cost and low-risk that reliably perform the required function fail to be obvious, and familiar courses of action that are obsolete are still among the first courses of action considered.

Some courses of action, then, have a prima facie considerability. That is to say, some courses of action are such that the burden of evidence is on the skeptic to deny first-considered status. One might wonder why one would bother casting doubt on some course of action before deliberation even begins. After all, deliberating about what should be deliberated on is itself deliberation, and if a course of action is brought up in the course of deliberation, then it is being treated as considerable.

Here, we should consider a spectrum of what we might call the research costs of deliberation. Often, when we begin to deliberate about possible courses of action, the possible consequences, constraints, and other relevant matters are fairly apparent. Other times, the courses of action themselves require inquiry, because the particulars of some context make it such that consequences, constraints, and other relevant details are unclear. Inquiry is required to do the deliberation of the courses of action justice. There is, on one side, then, cheap deliberation, and on the other, costly deliberation. When a course of action is recommendable, the burden is on the skeptic to show that the course of action should not even be investigated.

I have here been speaking of "courses of action." The topic of this essay is more theoretically oriented. Recommendability, as I have discussed, has a practical bent to it. Theoretical problems and issues require theories and the like. These may seem to be opposed to courses of action, but that is to understand courses of action narrowly.

Even in armchair philosophy, thinking, research, discussing, and writing are actions we engage in. Theorizing involves a course of action. That is not to say that theory and practice are not distinct spheres, but when we theorize, deliberate on theoretical issues, and inquire into theoretical questions, there are possible theories that are not worth consideration. When writing a philosophy paper, one does not consider every possible objection.

This becomes obvious when we consider that theories come in various degrees of fringiness. There are views that were once thought of as contenders that became fringe, and theories that were once fringe that became contenders. There are theories that seem to pass back and forth between fringehood and contenderhood. This at least seems to be the case in philosophy and science. As a rule, one addresses the contenders. Even when trying to move a theory from the fringe to contenderhood, we generally do not attend to other fringe theories, but to current contenders, and show that our favored theory is worthy of such a status in relation to them.

As in the case of feasibility, we can distinguish between theoretical and practical recommendability. Theoretic recommendability comes down to whether a theory or method should be among the first considered theories when addressing some theoretical problem or issue whether the theory or method stands in an approximately equal position among strong alternatives with respect to its theoretic merits and demerits.

In the case of this essay, I am defending the notion that conceptual engineering is a method that can be applied to even basic concepts. I want to show that the burden is on the skeptic to show why, in some particular case, a conceptual engineering project should not be considered. I count the thesis successful if I can establish that replacing *truth* is on the table and is at least a reasonable pursuit with respect to addressing the liar paradox. This can be generalized to other problems and issues concerning *truth* and other basic concepts.

2.2.1. The Objection from Centrality

Patrick Greenough (2014) argues that Scharp is engaged in what he calls "Conceptual Marxism." That is to say, the changes that Scharp is suggesting are very radical. He calls this Conceptual Marxism because implementing Scharp's suggestions would amount to

revolutionizing our conceptual repertoire. What I take to be Greenough's primary thesis is that if we replace *truth*, any concept in which *truth* plays a constitutive role will have to be replaced as well.

For example, if we take *knowledge* to be a justified true belief plus some anti-Gettier condition(s), then *knowledge* has to be replaced, because *truth* plays a central role in what it is to be *knowledge*. *Knowledge* is partly defined in terms of *truth*. If we are replacing *truth*, in at least philosophy, logic, and other related areas of inquiry, and *knowledge* is a target of philosophical inquiry, then *knowledge* has to be replaced here, barring some reason to restrict replacement.

This is what I call the objection from centrality, and it can be generalized to other basic concepts. A concept is central when it plays a constitutive role in many other concepts, and those concepts themselves play a constitutive role in other concepts, and so on. As far as analytic philosophy goes, *truth* is as central as they come, as *truth* arguably plays a role in constituting the following concepts: "provability, assertion, belief, ...inquiry, objectivity, reality, knowledge, judgment, evidence, justification, confirmation" (Greenough 2014, 11) among others.

Belief is also a central concept in analytic philosophy, playing a constitutive role in *knowledge* as well as *rationality.* It goes on to play their own constitutive roles in other concepts. If *truth* is unique then, it is unique in its degree of centrality.[7] This is the basic gist of the objection.

Greenough's specific objection and Scharp's replies are very centered on constitutive principles and the CIP. Greenough states:

> Just how radical is Scharp's replacement strategy? Very. Far more radical than is acknowledged in *Replacing Truth*. This is revealed when we consider the relationship between the concept of truth and other concepts such as provability, assertion, belief, knowledge, and more. Take the concept of (informal) proof and

[7] This is part of what makes it the ideal test case in considering the feasibility and recommendability of conceptual replacement for basic concepts.

> (informal) provability. With respect to these notions, Scharp recommends that we replace the… ...constitutive principle connecting provability and truth... (Greenough 2014, 9)

The constitutive principle to be replaced (T) and the proposed replacement (R) are as follows:

(T) If 'S' is provable, then 'S' is true

(R) If ‘S’ is provable [then] ‘S’ is ascending true.[8]

Greenough continues (10-11):

> Principle [T] is a constitutive principle for provability. To replace this principle with [R], while keeping the concept of provability unchanged, is to be engaged with conceptual revision and not conceptual replacement, as we have just seen. The Concept Identity Principle enforces the result that to replace [T] with [R] means that we are now dealing with a different concept of provability. Crucially, we are not allowed to use the word ‘provable’ to pick out this new concept...

One thing to immediately note is that, at least as far as I can tell, Scharp never demands a new word for every new concept. Yet, Scharp does not deny this claim in his response to Greenough. What can be said is that if what we mean by “word” is the specific sound and graphical representation of constituents of things like sentences, then, regardless of what Scharp thinks, a new concept does not require a new word. If "true" in "that sentence is true" and "her aim was true" is the same word in at least some sense of "word," then we can use the same word for different but closely related concepts.

Scharp's proposal does strongly warrant replacement words, because we are to replace “true” with two different concepts. While both concepts together to create a new semantics for the word “true” that works almost identically to truth as we know it, it will require new nouns or noun phrases just to distinguish the two concepts from each other. However, that will not be the case when one concept is replaced by one concept. A replacement word may be the better option

[8] I have replaced "with" with "then" here. I assume that's what was originally meant, but perhaps I have misunderstood.

in some cases, but surely not all. I understand Greenough's point as being that when we replace a concept, if that targeted concept helps constitute some other concept, then we have to replace that constitutive principle with a new constitutive principle.

The new constitutive principle would have the replacement concept take on the role of the targeted concept: this is what makes the new constitutive principle new. By the CIP, this makes for a new concept altogether, and this process can cascade as the newly created concept forces the creation of new concepts, which force the creation of new concepts, and so on. If at least one of *belief*'s constitutive principles is partly constituted by *truth,* then replacing *truth* causes us to replace *belief,* which causes us to replace *knowledge*, and so on.

By the end of Greenough's paper, he seems to think that the main problem is the CIP. However, if we were able to cast these changes as revisions rather than supplantations, it is not clear that it would make that big of a difference. The problem is that revising a concept creates the need to revise a bunch of others in a similarly cascading manner. Revision is still replacement in that one is still replacing one version of a concept with a new version, and so revising *truth* revises belief which revises *knowledge*.

As I argued in 1.3.1, what ultimately counts as revision or supplantation is going to be contentious, and there is no prima facie reason to think that any developed distinction is going to make it such that revisions are inherently less risky or less resource intensive. Ultimately, what the objection aims to show is that the costs of replacement are just too high, especially when it comes to central concepts (which unsurprisingly, have a lot of overlap with basic concepts), and as such, cannot be recommendable.

2.2.2. The Objection from Epistemic Loss

Someone might accept Greenough's general argument but think that this kind of implicit privilege given to conservation has no merit, or that whatever costs conceptual Marxism might have, the benefits are far greater, or at least can be in particular cases. There remains a somewhat related problem. Mona Simion has proposed a normative limitation centered on epistemic loss. Simion endorses the view that conceptual engineering should be given a wide berth as a tool to be used for moral and epistemic improvement, among others. However, she argues that it does run into a wrong-sorts-of-reasons problem. Here, a quick detour into what a wrong-kind-of-reasons problem is will be helpful. She sums up what a wrong-kind-of-reasons problem is quite nicely:

> In a nutshell, a reason is said to be 'of the wrong kind' when, although it counts as a consideration broadly in favor of phi-ing, it fails to bear on whether phi-ing is valuable. To say that something is a wrong kind of reason, however, is not to say that it is a bad reason: some reasons of the wrong kind seem to provide excellent support for phi-ing, while still failing to render phi-ing into a valuable action or attitude. (Simion 2018, 9)

One example would be desiring something one finds undesirable to avoid some kind of negative consequence. For example, making oneself desire mustard to avoid ridicule from a parent. All things considered, that is a reason to try and desire mustard but fails to make mustard desirable. Another good example of this is believing a falsehood for some kind of reward, like believing that New York is the capital of the USA for a million dollars. Getting a million dollars might be a good reason to try and believe that, but it does not make that a good belief. It is a paradigmatically bad belief, in fact.

The way this becomes a problem for conceptual engineering is that we might have a good reason to conceptually engineer something, but that reason fails to make the new or revised concept a good or even better concept. Simion paints a scenario. Take the concept *deer*. Let us

stipulate that *deer* does carve nature at its joints. A species of deer, roe deer, is as a whole not doing well. They lose 90% of their fawns to predation, while they lose significant numbers due to disease. As a result of this precarious position, roe falling under the concept *deer* puts the roe deer population at even greater risk "since, for instance, neither hunting nor protection legislation discriminates between roe deer and less vulnerable deer populations, roe deer is more likely to be hunted down, and less likely to be subject to protective measures." A case could be made for conceptually engineering *deer*.

The problem, according to Simion, is that the new concept would not be an epistemically successful concept; that is to say that whatever the new concept or concepts used to be, they would be worse concepts for failing to carve reality at its joints or do so in a way that is somehow worse than *deer*. This would count as an epistemic loss. Here is where Simion proposes what she calls:

> The Epistemic Limiting Procedure (ELP): A representational device should be ameliorated iff (1) There is all-things-considered reason to do so and (2) The amelioration does not translate into epistemic loss. (Simion 2018, 10)

The ELP is supposed to be a limitation of the scope of permissible conceptual engineering projects. How does this pose a problem for engineering truth?

As discussed in the last section, if replacing *truth* leads us to replace *knowledge*, then the case could be made that counts as an epistemic loss. In a world where *knowledge*, as a concept, is no longer in use, except in perhaps very niche ways, then can we be said to know anything? The question boils down to whether old concepts still apply when they are no longer in use. In general, I would suspect so. The older concept of *fish,* which applied to any vertebrate who lived in the water, still applies to whales, even though the contemporary concept of *fish*^{2} does not

apply. There is, however, a sense in which there is possible knowledge loss, and it is this sense that Simion is concerned with. If *deer* were replaced and fell out of usage, there would be knowledge lost, even though some people would know which critters *deer* applied to. We have to distinguish between collective knowledge and individual knowledge. There are tons of advanced physics, advanced biology, and other such things that WE know, but I definitely do not. If *deer* falls out of usage, then, per the example, we will be using an epistemically defective concept that helps roe deer but leaves us with less knowledge. This is epistemic loss, even though there would be people who still use and apply *deer*. Does this apply to *knowledge* though? I think perhaps especially so.

If knowing something *S* entails that one knows that they know it (the KK principle), then if knowledge gets replaced, then lots of people would outright fail to know things since they do not have *knowledge*. Unlike fish or deer, they would not count as knowing things, even from the perspective of someone who has the concept knowledge, because to count as knowing something, one has to have the concept knowledge to know that they know something *S*. They have to have a justified true belief that their belief in *S* is justified and true, and this relation constitutes knowledge to count knowing one knows *S*. Now, of course, the KK principle might not hold. I do not believe that it does hold! Nonetheless, it could very reasonably hold. Whether or not it does hold, people would nonetheless not know that they know anything because they do not have the concept *knowledge*, and hence, it would still count as an epistemic loss.

Some might take this to mean that we should reject the ELP. I am inclined to agree if we are treating ELP as fairly strict, where only the most exceptional cases get to bypass the ELP. Yet, I think that it does get something right. Knowledge is probably something valuable and losing it should count against a course of action, such as replacing a particular concept. If we

take the ELP as a principle of what should count against a particular case of conceptual replacement, then that is worth keeping. It could help explain the appeal of conservation, as well. This is one place where the distinction between revision and supplantation could make a normative difference. If a concept is only revised, then, arguably, there is no epistemic loss, at least, in a lot of cases. If we treat the ELP as a means of evaluating purposive conceptual change, then revisions have an advantage in certain cases, such as the case of replacing *knowledge*. If revising means that there is no epistemic loss, then revision has a strong advantage over supplantation. This is going to be contentious since, as mentioned before, it is going to depend on the correct account of the nature of concepts and their identity. If the CIP is right, then almost any kind of conceptual change is going to count as supplantation. Anything less strict than the CIP is going to vary as to what counts as a revision and a supplantion. This leaves us a bit stuck, as we have what seems to be a reasonable evaluative principle and no way of telling whether we are implementing it correctly, given the open questions regarding the nature of concepts. What we need is a way around these issues. Ultimately, the objection is supposed to show that the ELP forbids replacing *truth* because of the epistemic loss that results from replacement, and the lack of dire consequences to justify that loss. Even on a weaker conception of the ELP, it would work very strongly against the recommendability of replacement, because the epistemic loss caused by replacement would be too much to overcome from other considerations.

2.2.3. The Objection from Underjustification

The objection from lack of justification can be summed up by echoing an old adage: if it ain't that broke, it is not worth fixing. The simple form of the argument goes something like the following. We use the concept of *truth* in matters practical and theoretical all the time, and usually without much complication. Whatever problems liar reasoning gives the concept *truth*,

they do not seem to affect our ability to use *truth* in contexts where liar reasoning is unlikely to appear. Liar reasoning is unlikely to appear in most contexts. If there were a need to replace *truth*, then there would be significant problems in using *truth* in addition to the problems directly posed by the liar. There are no such significant problems. Therefore, there is no need to replace *truth*.

We can tweak an argument provided by Eklund (2014) to give us a more sophisticated version of the argument above. As I understand Eklund, there are at least two justifications for replacing *truth*. I will quickly summarize each before explicating each more deeply. One motivation would be to adopt the *truth*-like concept that our minds tacitly use, insofar as it is a different concept from the ordinary concept of *truth.* That is to say that if the concept that we employ internally is different than the one that we talk about, we could replace the latter with the former. That would create motivation to replace *truth* because it accords with our goal of giving us a more accurate and scientific picture of the world. Second, if using *truth* gave us inconsistent beliefs, that would create a need to replace *truth,* because having inconsistent beliefs is generally thought to be a significant defect.

The first justification Eklund calls the cognitivist justification. It is ultimately the same justification that is generally standard for truth-theoretic semantics:

> People have tacit knowledge of the referents of words. They know facts such as that "Newt Gingrich" refers to Newt Gingrich, "dog" refers to dogs, etc. They know the syntactic structure of sentences, and they know the composition rules corresponding to the syntactic structures. This lexical and compositional knowledge allows them to grasp the truth conditions of sentences they hear, and express propositions they want to express. Truth-conditional semantics states what speakers have tacit knowledge of, and thus provides a psychological explanation of how our knowledge of language allows us to gather information about the world, and communicate it as well... The cognitivist can be seen as treating semantics as a branch of psychology. (Eklund, 2014, 296)

By this line of reasoning, any concept that we employ in semantics must be one that humans have tacit knowledge of, whether that be truth or its replacement. One might think that truth-theoretic semantics always employs truth, resting on the assumption that our normal concept of *truth* is the one we use and have tacit knowledge of, and thus, any replacement would fail to get off the ground. Eklund points out that we may have tacit knowledge of both our ordinary concept *truth* and some other concept used for semantic processing (understanding the meaning of words, sentences, etc.), but the problem is that it is merely a possibility. More is needed to pursue replacement. Another problem is that a kind of psychological realism about tacit knowledge is needed for the cognitivist justification to work, and psychological realism lies on shaky ground. "...[A] natural objection to the cognitivist conception is that it is psychologically implausible. Do really all language users, including, for example, very young children, possess a concept of truth?" (Eklund 2014, 296).

The other justification we can call the doxastic justification. Again, the idea is that if the properties of the concept of truth lead to inconsistent beliefs, we want to replace truth with a concept that does not lead to inconsistent beliefs, because having inconsistent beliefs is irrational, perhaps paradigmatically so. We can think of the thesis of truth having such properties as characterizing truth's inconsistency, and inconsistency of concepts in general. More specifically, it is a view on exactly what sort of things are inconsistent with one another when we become competent with an inconsistent concept: when I know how to use a concept, when I understand a concept, in virtue of what does the inconsistency manifest, what are the things that contradict one another? The belief view maintains that a particular concept is inconsistent in virtue of the inconsistent beliefs produced by said concept. However, characterizing conceptual inconsistency in terms of the inconsistency of beliefs is only one option, and it is not an

especially strong one, relatively speaking. More specifically, it is a view on exactly what sort of things are inconsistent with one another when we become competent with an inconsistent concept: when I know how to use a concept, when I understand a concept, in virtue of what does the inconsistency manifest, what are the things that contradict one another? The belief view maintains that a particular concept is inconsistent in virtue of the inconsistent beliefs produced by said concept. The alternatives to the belief view can actually accept the first possibility. That view that rules, dispositions, and entitlements can come in different strengths or rankings is relatively non-contentious. A more basic entitlement to deny contradictions can exist, along with an entitlement to inconsistent claims. A disposition to accept inconsistent claims can exist, along with stronger dispositions to reject that which is inconsistent. Rules that override other rules are common in systems of rules. More importantly, these things not only exist along with one another, but they can also be acknowledged by the people under their sway. It is not obvious that this is the case with belief. However, characterizing conceptual inconsistency in terms of the inconsistency of beliefs is only one option, and it is not an especially strong one, relatively speaking.

What are our other options? Eklund outlines three other views: the dispositions view, the normative view, and the psychological view. The dispositions view is that competence with an inconsistent concept gives one a disposition to accept inconsistent claims. The normative view is going to take competence as normative, and the notion of inconsistency will have a normative flavor. Scharp's view, for example, is that competence with a concept entitles you to claims, and an inconsistent concept entitles you to inconsistent claims. Finally, under the psychological view, being competent with an inconsistent concept means having internal semantic processing rules from which contradictions can be derived.

What all three views have in common is that they would not entail inconsistent beliefs. This becomes clearer when I explain how one can have multiple dispositions, entitlements, and rules at the same time. Ultimately, the argument comes down to the following. First, as far as inconsistency goes, only the inherent irrationality of holding inconsistent beliefs is enough to justify replacing a concept. Second, three of the four ways we might under conceptual inconsistency do not entail inconsistent beliefs. Third, the one that does (the belief view) is not promising. If the belief view is unlikely to hold and the other inconsistency views do not entail inconsistent beliefs, then it is unlikely that inconsistent concepts entail inconsistent beliefs. Hence, the justification for replacing a concept must come from somewhere other than its failure to be consistent.

One might challenge the claim that only inconsistent beliefs (with respect to a concept being inconsistent) justify replacement. One might also challenge the claim that the belief view is unlikely to hold. We can meet both challenges in explaining the so-called expert objection. The expert objection is that when someone is an expert with an inconsistent concept, they do not cease to be an expert when they determine that a concept is inconsistent. One can make this objection against all four views. To illustrate, let us consider the expert objection against the belief view with respect to Okie. If I am competent with Okie, I will apply it to people from Oklahoma but not to non-scummy people. As discussed before, Okie entails the non-existence of non-scummy Oklahomans. According to the belief view, my competence with Okie leads me to believe there are no non-scummy Oklahomans. There are non-scummy Oklahomans. When I realize I have inconsistent beliefs as a result of my competence with Okie, I reject the belief in the non-existence of non-scummy Oklahomans, I should lose my competence with Okie. This seems implausible. Despite rejecting the belief in the non-existence of non-scummy

Oklahomans, we would still understand the implication of "We don't want Okies around here." On the other views, if I lose my disposition, entitlement, or processing rule, I should lose my competence (this will turn out to be moot). Yet, anybody reading this should be competent with Okie. There seem to be two possibilities. One, we have not lost the inconsistent beliefs, processing rules, entitlements, or dispositions. Two, our competence with inconsistent concepts is an illusion of some kind.

Beliefs might come in degrees or strengths, perhaps even along more than one axis. I had put money on it. However, this is not the issue. Beliefs are one of the primary targets of rational evaluation, if not the primary target. It is often thought, not without contention, that the theoretical success of an individual stands and falls on that individual's beliefs. There are many elements involved in theoretical success, but insofar as our theoretical pursuits are ultimately about the attainment of knowledge, then it is beliefs that take center stage[9]. Nearly without exception until fairly recently in the Western tradition, holding contradictory beliefs was seen as a serious theoretical defect, and it is still by far and large the predominant view. Inconsistent dispositions, inconsistent mental processing rules, and other possible ways of understanding inconsistent concepts, simply do not carry the same weight of rational failure. One might be less rational for having a disposition to accept inconsistent claims relative to someone who does not have that disposition, but one is not being irrational merely having that disposition. This is analogous to desire and action: we might have a desire to act irrationally, and maybe that makes us less rational than someone who does not have that desire, but unless we act on that desire, we

[9] Knowledge-first accounts complicate this picture. However, the JTB+ is still more than widespread enough to motivate the objection.

are not being irrational. The problem for belief is that holding contradictory beliefs is irrational,[10] and that means that on the belief view, we are either competent at the cost of being irrational, or under the illusion of competence. This would mean that learning an inconsistent concept would doom you to inescapable irrationality, by contradictory beliefs or living under an illusion. This is a good reason to prefer one of the other views.

What I hope to have shown with the expert objection is that the other views can tolerate inconsistency and competence together in a way that belief cannot, because of the central role that belief plays in evaluating rationality. In the course of showing this, I explained why beliefs are central in that way, and this explains that inconsistent beliefs are going to justify replacing truth in a way inconsistent dispositions and the like are not. Either inconsistent beliefs are not brought about by truth being an inconsistent concept as required by the doxastic justification, or the psychological realism required by the cognitive justification is implausible. The idea is that we lack sufficient justification to replace truth (and perhaps other concepts as well). Of course, there could be other justifications, like the ones presented in the first chapter, but those justifications depend on truth's inconsistency in itself being enough to justify replacement.

[10] On views that involve degrees of belief or credence, we would need a different account of irrationality. I'm using the more traditional notion of belief for clarity sake.

A Framework for Addressing the Objections

3.0.0. The Cognitive Role of Truth and its Replacements

In this chapter, I engage in some conceptual engineering myself and explore a possible framework that utilizes what I call cognitive roles and cognitive tools, and I show how it can fully address some of the issues raised in Chapter Two and at least partially address the rest. By giving examples and making some arguments, I want to show that this is a plausible framework, but it is not to be taken as something established. I make no serious attempt to establish it here. My hope is that the value of the framework comes through its intuitiveness and its ability to address the issues in Chapter Two. I start by giving a sketch of what cognitive roles and tools are and how they work, and what the theory is supposed to accomplish. I then establish plausibility by arguing both that the concepts of *cognitive role* and *cognitive tool* at least approximate some genuine phenomena and that they provide a framework for thinking normatively about concepts. I further argue that, even if the previous arguments fail, the concepts give us a good framework for thinking about conceptual engineering. Finally, I show how cognitive roles address the various objections brought up in Chapter Two.

3.1.0. The Cognitive Toolkit Framework (CTF)

In this section, I go into detail about what cognitive tools and cognitive roles are supposed to be and how they are supposed to work. I use both mundane and abstract examples to show how they apply to day-to-day life and theorizing, including how cognitive tools and roles can work expressively. These examples are meant to explain but also give a sense of plausibility to the concepts and the framework as a whole. With respect to the theoretical, I use examples from both ethics and epistemology. I use this to more fully introduce conceptual roles. I give a

short account of the notion of problems and show how cognitive tools and roles are grounded in problems.

3.1.1. Cognitive Tools and Roles

The best place to start is to explain what a cognitive tool is supposed to be. According to my potential framework, a cognitive tool is just anything that helps one to think, broadly construed: symbols, words, memories, colors, notepads, concepts, stories, thermometers, feelings, fingers, etc. A cognitive tool helps one think by fulfilling a role in one's thinking, a cognitive role. It might be obvious already that I am using "cognitive" here in a pretty broad way: "cognitive" here just means involved with thinking, loosely construed. It is not necessarily exclusive to representational or symbolic reasoning. Other kinds of thinking or activities involving thought could include things like self-examination, self-therapy, emotional regulation, carefully working with one's hands, game-playing, decision-making, articulating desire, and so on. These would all be different kinds of thinking or activities that involve thinking in my sense of "thinking."

Let us start with something representational in nature. Consider the concept of *credit card* at the level of the individual. *Credit card* allows me to recognize tokens of credit cards, allows me to understand the word "credit card" used in a sentence, and allows me to empathize with the alarm someone has when they believe their credit card has been stolen. I can use *credit card* to help decide whether I want a credit card or to use a credit card, rather than a debit card or cash. *Credit card* helps guide my inferences, imagination, and actions concerning credit cards. There is also an aspect that is not purely linguistic that we might call the image involved in *credit card*. It is the vague idea of what a credit card looks and feels like: a general sensory picture of the object type. If a friend asks me to grab their credit card off the table, I have a sense of what the credit

card looks like, so that when I cast my eyes over the indicated table, I can recognize the right object as a credit card, regardless of whether I have ever seen said credit card. If the credit card is misplaced, and I am searching the couch by sliding my hands through the cushions, I have a sense of what a credit card feels like in my hands, so that if my hand comes in contact with it, I recognize it as the thing I am looking for. Most likely, people from a society without card-like payment methods would not have the concept of *credit card* and would likely have a difficult time thinking about credit cards. They would have a harder time retrieving a credit card without having that vague image associated with *credit card*. Even in the case where they knew that "credit card" referred to small rectangular pieces of plastic but this was all that they knew, it might be difficult to empathize with the level of panic of someone who exclaims, "I can't find my credit card!" *Credit card* is the cognitive tool that makes it possible for the individual who possesses *credit card* to interact with and think about credit cards qua credit cards. For the individual, *credit card* fulfills the credit card role.

Cognitive tools, as I conceive them, can also play what we might call expressive or imperative roles. The color red, in certain contexts, can express or implore caution or the need to stop. It can demand attention. Consider a stoplight or the red warning label. Consider the red notification symbol. The way red is used in these contexts is less about describing the world, and more about getting people to behave in a certain way. Video game design often tries to take advantage of things like this. In a two-dimensional, pixel-based video game, pixels arranged in a triangle-like shape are often a way of communicating to the player that interacting with these objects in the game world will cause death or damage to the player character, i.e. cause the player to lose the game, or otherwise diminish the player's success. Red and the pixelated

depiction of spikes both can act as cognitive tools in the cognitive role of expressing importance, danger, and virtual danger.

The examples so far have been fairly concrete. These are cognitive tools that fulfill fairly practical roles where the upshots are straightforwardly action-like in character. We might wonder if more abstract and theoretical notions fit neatly in the conceptual toolkit framework. Having described what I take to be intuitive examples, I will move on to examples from two branches of philosophy: ethics and epistemology. These examples, I hope, will show that the cognitive role theory works well for the abstract and theoretical, as well as allow me to better discuss the nature of cognitive roles[1]. I will start with an example from ethics. Consider two, overly simplified, consequentialist theories of ethics: a vulgar Benthamite utilitarianism where happiness (pleasure minus pain) is to be maximized, and a consequentialism where freedom is to be maximized that we will call "libertarianism." In both cases, we have a maximizee, the quality to be maximized in our conduct. Both happiness and freedom are cognitive tools that fulfill a cognitive role. To see how, note that both utilitarianism and libertarianism fulfill the cognitive role of an ethical theory. At least one cognitive role that any ethical theory plays is the role of determining future behavior, insofar as controlled thought influences that behavior. When I adopt an ethical theory, ideally, I think about which behaviors and patterns of behavior align with that ethical theory and I attempt, through thought, to influence myself to behave in those ways. Consequentialism fulfills this role by directing our intended behavior toward the consequences, and utilitarianism and libertarianism fulfill a more specific role by directing said behavior toward particular kinds

[1] I am not remotely suggesting that my two examples are *merely* cognitive tools/roles. At most I am suggesting that are cognitive tools and fill cognitive roles, whatever else they might be.

of consequences. When conventionally straightforward[2] with respect to my ethical beliefs, if I am utilitarian, I organize my thinking and actions around happiness, and if I am a libertarian, freedom (*happiness* and *freedom*).

With regard to epistemology, we can look at Gettierology. One way of framing the goal of Gettierology is to find a concept of knowledge that is not susceptible to being attained by dumb luck. In this vein, many tried to come up with various anti-luck criteria for knowledge. One proposed criterion is to add an infallibility condition to knowledge, so that knowledge must be an **infallible** true justified belief. Another is to disallow any false evidence. We can say that these serve the anti-luck role. Insofar as knowledge is supposed to be an achievement, attaining true belief and even justified true belief by mere luck undermines that achievement. We can frame Gettierology as the collective effort to pick out a class of justified true beliefs worthy of the role of knowledge: one class of beliefs that count as cognitive successes. The chosen anti-luck condition is the tool that we use to pick out that class, so we have an anti-luck condition embedded within a knowledge role, which is further embedded in the role of a distinctive kind cognitive success. We can see here that cognitive roles are often also cognitive tools themselves. The knowledge role is the cognitive tool that plays the role of cognitive success but so could understanding or mere true belief.

---

[2] I say "conventionally straightforward" here because, at least with act consequentialist theories, ethical theories are treated as if they as we consciously guide our behavior to conform with our adopted theory. For example, I have to consider which act will produce the ethically correct outcome if I'm some kind of act consequentialist. Even with something arguably less directly intellectual like a virtue theory, I still have to be vigilant with respect to my behavior and whether it is aligning with the virtues I am supposed to be embodying and training myself into the correct virtue when there is a pattern of behavior that does not align with the virtues. There are different approaches, however. Railton has proposed that one adopt one theory methodologically, so to speak, while believing some other theory to be the correct theory (Railton 1984). I personally inclined to think that only by adopting a virtue theory behaviorally will you bring about the best consequences.

3.1.2. Problems

We might wonder then whether it's cognitive roles all the way down. While I do not want to rule out this possibility in some cases, generally cognitive roles are formed by problems. Why problems? Every human discourse, every field, every discipline, every project, deals with problems. Arguably, even non-human life deals with problems. Thought itself can be said to have come about to confront problems. Basing cognitive roles in problems gives the framework a very wide application. What is a problem? A whole theory of problems would take this thesis too far afield, but we can give a sketch of what a problem is. A problem arises any time there is an obstacle in accomplishing some function or goal, something in the vein of "I cannot (easily) accomplish **x** because of **y**." Another form problems take is the form of "My goal is **x,** but I have a sub-goal **w** of accomplishing **x** under constraint ***z*** and **y** prevents me." In both cases, **y** is the problem, but only in virtue of preventing the goal or sub-goal from being accomplished. There must be some goal that is obstructed in some substantive sense due to **y.** A case of the former might be something like "My goal is to make a basket in basketball, but I am not strong enough to get the ball to the proper height," while an example of the latter would be something like "I want to run 100 yards in 16 seconds, but I am not yet fast enough."

With respect to function, some problems don't have any need for cognitive roles, because some problems don't require thinking. Given that most creatures have the goal of staying alive, the immune system fighting off a disease solves a problem without that critter thinking about it. Other problems do require thought to be addressed. Cognitive roles are formed by the demands of a problem. Thoughts can be constituted such that they can meet these demands; these are the cognitive tools. Why have the middleman of cognitive roles when we could just have cognitive tools and problems? A cognitive role is a mental articulation of what the demands of the problem

are. For example, a mind determining for itself what is broadly needed to address a problem; if I think about what I need to solve the problem of driving nails into wood, then I would look for something with a flat surface, hard, relatively heavy but liftable, and also graspable. When it comes to an anti-Gettier condition, we need something that eliminates the possibility of knowing luckily while avoiding an overly high burden that results in skepticism. In both cases, there are potential tools that can fulfill those roles. A problem can, and often does, extend beyond the way its demands are conceived. Some problems can be understood in different ways. That lets us form a different cognitive role.

A cognitive tool successfully fulfills a cognitive role insofar as it addresses the problem articulated by the role. There are at least three ways to address a problem: solution, dissolution, and domestication. Remember, what a cognitive role does is articulate the demands of a problem, and the demands are themselves the articulations of which properties are required of a tool to achieve the goal. A solution is a tool that has the properties required for achieving the goal. A hammer has the properties of being mobile and having an unobstructed flat hard surface. A solution to the Gettier problem would minimally be a condition of knowledge with some property that would, in principle, prevent the possibility of lucky knowledge. A dissolution works differently: it addresses a problem by somehow bypassing the demands. There are likely many ways of doing this, but one is rearticulating the problem at hand or some problem further up in what we might call "the problem hierarchy." Dissolving is a matter of rearticulating a problem in a way that removes demands such that it resolves a problem, either that problem, or a problem further down the hierarchy. A business might dissolve the problem of choosing a distributor by selling directly to its customers. The Gettier problem could be dissolved if it could be shown that a concept of knowledge that allows for lucky knowledge was defensible because

an anti-luck condition is no longer necessary. Domestication is making a problem more manageable. The problem of cleaning is a problem that can generally only be domesticated. A house doesn't stop needing to be cleaned because it continuously becomes dirty. Cleaners and tools can make cleaning easier and require less work. We could think of conceptual inconsistency as a problem that can only be domesticated. We can solve particular instances of inconsistency, but inconsistency will rise again. We can develop strategies that make dealing with inconsistency easier, such as the conceptual fission that Scharp employs with truth's inconsistency.

3.2.0. Plausibility

Here, I will offer some arguments in favor of both the descriptive and prescriptive aspects of the account. These aspects, strictly speaking, are free-standing. If I am wrong about how and why thoughts are implicitly used, that says very little about how they should be used. The descriptive aspect may support the prescriptive aspect, insofar as the account approximates how and why we use thoughts at least implicitly, it will probably be easier to do so in a self-aware way. That is to say that improving on something we already do is less normatively onerous than having to deeply reform our behavior.

I'll begin with the descriptive aspect. I only aim to convey how we seem to think about the different parts and aspects of thought in carrying out our work, theoretical and practical. When it comes to words, the appearance of synonymy can be explained as using different words to serve the same cognitive role. When we speak of a dog, we can use "mutt" or "mongrel" to indicate that a particular dog has multiple, unknown ancestors of different breeds. Whether "mongrel" or "mutt" actually mean the same thing is irrelevant, because my account is not an account of meaning, per se. If they do invoke the same concept, or at least invoke overlapping

sets of concepts, we can reasonably say that they both serve the role of invoking that overlap (the intersection of the sets)[3]. Of course, someone might point out that “mongrel” carries a more negative connotation than “mutt.” That strikes me as correct. However, there are all sorts of ways in which that negativity can be nullified, such as saying “mongrel” in a loving tone. However, even with the more negative connotation, “mongrel” works as well as “mutt” in so far as you are trying to indicate a dog that is of mixed breed. Compare with someone who uses “mule” to indicate that a dog is of mixed breed. “Mule” is used to indicate the hybrid offspring of horses and donkeys. The intended listener might understand the intended meaning but is less likely to, because although “mule” is used to indicate a hybrid, it is not used to indicate hybridness of breeds in dogs. “Mule” could work, but since it is significantly less likely to, it is not a suitable word to use.

There are at least some cases in the act of translation where the concepts being invoked by the respective words are not quite the same. Consider the English “fun” and the French “amusement.” My understanding is that “amusement” is often used to translate “fun” into French. However, “fun” and “amusement” are not invoking the same concept, each of them calling in the use *of fun* and *amusement*. At least it was so at one time, as “fun” has been said to have no proper counterpart in other languages[4]. They do not have the same extension. There are,

---

[3] This could maybe be a way of explaining synonymy without appealing to the equivalence of meaning: the reason that we can use "mutt" and "mongrel" interchangeably in dog-related contexts is that they work just as well as each other.

[4] "Now this last-named element, the fun of playing, resists all analysis, all logical interpretation. As a concept, it cannot be reduced to any other mental category. No other modern language known to me has the exact equivalent of the English "fun". The Dutch "aardigkeit" perhaps comes nearest to it (derived from "aard" which means the same as "Art" and "Wesen" in German, and thus evidence, perhaps, that the matter cannot be reduced further). We may note in passing that "fun" in its current usage is of rather recent origin. French, oddly enough, has no corresponding term at all; German half makes up for it by "Spass" and "Witz" together. Nevertheless it is precisely this fun-element that characterizes the essence of play." (Huizinga,

for example, very difficult video games that are undoubtedly fun (for at least some people) that would not be properly described by the French “amusement.” There is little joy or mirth in these games. They are frustrating, maddening even. There is, at best, the thrill of hard-earned victory. They are fun, nevertheless. The pleasure comes from the challenge. Some films are accurately described by “amusement” but aren’t “fun.” Nonetheless, “amusement” is a fine translation where *fun* and *amusement* overlap, or insofar as there is an intended indication that some type of pleasure will be had upon consumption of said media. There are other possible translations, but what translation we use is determined by the goal of communicating what is being translated in the new language.

In both cases, what matters is that we achieve a certain communication goal, and, in each case, there are plenty of contexts in which the communicative goals can be achieved by either of the respective options. In the “mutt/mongrel” case, being words of the same language with identical or nearly identical extensions, are almost always going to be able to serve the same cognitive role. The “fun/amusement” case, on the other hand, is going to have a narrower overlap. There are going to be only certain classes of cases where you can translate “fun” as “amusement” and vice versa. Yet, those are cases where they serve the same cognitive role.

Again, the cognitive role theory does not only apply to linguistic and language-adjacent mental entities. The framework is flexible enough that anything that might play a role in thought could work, including both physical objects and external representations. I return to video games here. In addition to the depictions of spikes, game designers will also use depictions of fire/lava

---

1938, 3) This is likely premature both for reasons of linguistic change and reasons more pernicious. It is at least possible that "amusement" does sometimes invoke *fun* and it is difficult to say what languages Huizinga might be leaving out due to something like Eurocentric bias. That said, the French speakers I have spoken to have agreed to the notion that “amusement” does not mean “fun,” at least not quite.

and electricity to indicate danger and damage. These, along with spikes, are also more specifically used to indicate an obstacle. Given the way most video games work, what the designer chooses in a given design, whether fire, electricity, or spikes, makes no fundamental difference in the gameplay. Depictions of fire and electricity might have additional effects, such as continuous damage or rendering the player character immobile temporarily, but typically how one avoids such obstacles will be the same. These depictions can be equally effective at communicating danger and obstacles. In that respect, all serve the same cognitive role.

Let's do one more example, something that shows different cognitive tools (such as linguistic tools and non-linguistic tools) fulfilling the same role. Sarah is trying to find the engineering building in an area that she is somewhat familiar with. Sarah can memorize things in a very accurate and detailed way given a few minutes of willed effort. There are three scenarios. In (Des), she has memorized a thorough description of where the building is in relation to other features of the area and what the building looks like. In (Map), she has memorized a wordless map of the area with a symbol of where the engineering building stands. In (Img), she has memorized a high-quality aerial photo of the area with the knowledge of what the top of the engineering building looks like[5]. In all three cases, what she's memorized allows her to more easily find the building she's looking for. All three methods function at least somewhat differently. (Des) works primarily through linguistic ability. (Img) works primarily through visual memory. (Map) works through a combination of what might be considered a more general symbolic ability and visual memory. They are distinct tools, but they perform the same role in different ways. It is not difficult to imagine degraded versions of these examples that fail to work as effectively.

[5] We might say she remembers being on top of the engineering building roof.

Insofar as these examples illustrate how easy it is to think about concepts and other mental entities as fulfilling the same and different jobs, they suggest that my theory articulates the way we already think about how we should use words, concepts, and other entities. Next, I will argue that, regardless of whether my theory approximates the world, the framework could plausibly work to ground conceptual normativity.

One problem a theory of conceptual normativity has to deal with is that there are quite a number of different theories of what concepts are and how they work. One benefit of my theory is that it can be broadly neutral on what concepts are, because a cognitive tool can be anything that can be used in a tool-like way in thought. Even if it turns out to be the case that concepts are not mental entities at all (perhaps they are abstract objects), they nonetheless would still have some kind of mental counterpart that we employ for connecting to or grasping the concept. That's unlikely to make a substantial difference. What concepts turn out to be is very unlikely to make any difference to the way that my theory treats concepts or change the normative consequences involving them given that we employ concepts without any definitive metaphysical understanding already.

Grounding normativity in use or purpose is fairly intuitive and has a heritage going back at least to Aristotle. The notion that concepts solve problems is particularly intuitive. Everybody understands what it is to encounter a problem and what it's like to find the right sort of tool to solve said problem. This kind of familiarity with problems makes it so that just about anyone can evaluate concepts. This makes the theory relatively democratic as opposed to being a more esoteric theory.

The third benefit is that the theory is compatible with other theories of normativity, with different sets of values, and with a wide variety of ends. What constitutes a problem is going at

least be greatly influenced by one's values and notions of normativity, whether implicit or explicit. A subjectivist or a relativist is going to see problems as largely the function of an individual or particular group. Problems for the utilitarian will often be framed by maximizing pleasure and minimizing pain. Other problems will be framed in terms of epistemic concerns and goals.

3.3.0 The Conceptual Role of Truth

Now having explained what cognitive roles are, we can start thinking about the cognitive role of *truth*. That is the alethic role or roles as the case may be. That is, we can start thinking about the role that *truth* fulfills in our cognitive lives, about what problems are articulated in that cognitive role, and how truth fits that role. We can do this by working backward from truth to the cognitive role *truth* plays by asking what the concept of *truth* does for us; what it allows us to do. We can then examine why we want to be able to do these things by determining what problems these abilities address.

We can start with the kind of utility that deflationists point at when theorizing about *truth*. Deflationists want to strip *truth* of any metaphysical baggage, but to avoid rendering what was supposed to be a central concept totally useless, they often point to the ways that truth plays an important role in our discourse. First, *truth* is often used as a general device of endorsement and rejection (Scharp 2013, 63). We might say "That's true!" to agree with someone, or we might utter "I don't think that's true" to reject someone's claim. Another way *truth* is used is as a prosentence operator. A prosentence is like a pronoun, except whereas a pronoun can be used to substitute for a person's name, a prosentence substitutes for a whole sentence. We might use "she" instead of "Sally" and we might use "that's true" instead of "the bread is old." Finally, *truth* can be used as a device of generalization. One might say something like "everything James

writes in the *Principles of Psychology* is true" or "what Sally said yesterday is true." While these uses of *truth* are emphasized by deflationists, it does seem that it is used this way. If things are as they seem, these are part of the general alethic role.

For the more substantial aspects of *truth*, we can look at Canberra Plan-style platitudes about *truth* (Scharp 2021, 652). The Canberra Plan is a methodology of conceptual analysis in which one collects platitudes about some concept and tries to determine whether there is anything that fits those platitudes where a platitude is just a seemingly obvious statement that is seen as widely accepted. We can take Michael Lynch's list of platitudes about *truth* as our working example. Lynch has a list of six platitudes:

> (Objectivity) The belief that p is true if, and only if, with respect to the belief that p, things are as they are believed to be.
>
> (BS) The belief that p is true if and only if p.
>
> (TS) The proposition that p is true if, and only if, p.
>
> (Warrant Independence) Some beliefs can be true but not warranted and some can be warranted without being true
>
> (Norm of Belief) It is prima facie correct to believe that p if and only if the proposition that p is true.
>
> (End of Inquiry) Other things being equal, true beliefs are a worthy goal of inquiry. (Scharp 2021, 652)

We can think of at least some of these as describing an important function of *truth*. Consider (Objectivity). Under a representationalist framework, something true is supposed to represent the world the way it really is: the belief is supposed to somehow mirror, reflect, correspond to, or be isomorphic to the world. (Norm of Belief) says that beliefs that meet the criteria of (Objectivity) aare of the class of correct beliefs. Under this framework, *truth* serves the role of picking out a certain class of beliefs and acting as the standard of correctness for belief: it is this class selection

that makes it possible to serve as that standard. It is the dichotomy between true and not-true beliefs that serves as the basis for correct and incorrect beliefs respectively. This gives guidance when conducting ourselves epistemically; it gives us a goal in managing our beliefs (End of Inquiry). The rest of the platitudes clarify the relationship between *truth* and either *belief* or *proposition*. These platitudes in this way clarify the role of *truth*. For example, (Warrant Independence) distinguishes warranted beliefs and true beliefs. This does something like separating which beliefs are acceptable to hold and the ultimate correctness and goal of belief.

What problem does this role address? Under a representationalist framework, our beliefs are the primary way we represent the world. Insofar as representing well is possible and desirable, we want some criteria for representing well. *Truth* serves as this criterion with respect to beliefs. There are questions about why we should want to represent well, about whether representing well solves some problem or is valuable in itself. Even (Warrant Independence) might suggest a tension. If my take on (Warrant Independence) from above is correct, then the class of beliefs designated as acceptable to hold is just the class of warranted beliefs, and those beliefs may or may not be true. We might wonder whether *truth* is necessary when something like *justification* or *warrant* might be sufficient: why aim for true beliefs when warranted beliefs will do?

In responding to Richard Rorty on the uselessness of *truth*, Huw Price argues that *truth* has a function that *justification* and *warrant* cannot serve. *Truth* impels us toward agreement. To show this, Price asks us to consider three norms of assertion:

> (Subjective assertibility) A speaker is incorrect to assert that p if she does not believe that p; to assert that p in these circumstances provides prima facie grounds for censure, or disapprobation.

> (Personal warranted assertibility) A speaker is incorrect to assert that p if she does not have adequate (personal) grounds for believing that p; to assert that p in these circumstances provides prima facie grounds for censure.
>
> (Truth) If not-p, then it is incorrect to assert that p; if not-p, there are prima facie grounds for censure of an assertion that p. (Price 2003, 173-175)

He asks us to imagine a society that lacks the third norm, the Mo'ans. The Mo'ans will criticize for asserting things they do not believe and for asserting things they lack sufficient evidence for. If two Mo'ans disagree, they will evaluate each other for sincerity, as well as their evidence and reasoning. If they determine, however, that the other is sincere and justified or warranted in their belief, they simply accept the disagreement.

Without (Truth), there is no normative ground to improve their belief. There is no conceptual space provided for improvement; warranted beliefs are the best you can do. For Price, (Truth) provides the conceptual space to improve on one's belief in what Price calls "passive account," where the passive account merely makes improvement possible. Price argues (Truth) does not merely account for the possibility of improvement but actively encourages it i.e. provides an active account. Approval for settling disagreement and disapproval for ongoing disagreement adds oomph on top of the mere possibility of improvement. We can call this particular subrole "the conciliation role." Anything that plays the alethic role will also need to play the (Conciliation) role.

Having laid out the theory of cognitive roles and the role that truth plays, we are finally in the position to address the six objections brought up in Chapter Two. I will use the cognitive role theory to show that if there are concepts that can successfully fulfill the alethic role described above, then the objections made can be successfully refuted, or at least made less worrisome. I will begin with the objection from semantic externalism and proceed to these in the order they were presented in Chapter Two.

3.4.0 Addressing the Objection from Semantic Externalism

Here, I will address the objection from semantic externalism. I start by addressing the depth and scope of the problem before turning toward how we might generally approach the problem from the perspective of the conceptual role theory. I then present two examples of what seem to me to be successful cases of conceptual engineering that happened mostly without explicit institutional pressure, where human beings chose to start making using words that had named related concepts where those efforts have been rewarded. I then address Herman Cappelen's specific objections against the possibility of conceptual engineering.

3.4.1. The Scope of the Issue and the Conceptual Role Approach

The objection from semantic externalism is, in my evaluation, the hardest objection to answer. Here the beast may be wounded, but it is unlikely to be slain. Part of the difficulty is that, as it is, semantic externalism is a broad thesis. The extralinguistic anchors that fix words and concepts vary from one particular theory to the next. It would be difficult to address all the kinds of anchors: physical, biological, social, and so on. Another issue is that there is no prima facie reason to think that anchors equally fix all words and concepts. Perhaps some classes of concepts are unchangeably fixed and some are more open to modification.

Despite these issues, there is reason to be hopeful. The hope lies in the way that problems appear for us and the way we deal with those problems. In the theory illustrated above, it is problems that ground cognitive roles and tools. They are determined by the problem and judged as good or bad insofar as they address the problem. Problems arise out of our interactions with the world, whether those specific interactions are with ourselves, others, or the environment. We form goals or goals are formed within us, and often obstacles are formed in the same stroke since obstacles are obstacles in virtue of the goal they prevent us from achieving.

When we successfully address a problem, the way we think or how we interact with the world changes, usually in something like a positive way. On the other hand, real problems left unaddressed cannot be ignored without ill effect. At best, they can be given up on. That is to say that, when we are deprived of a means to our goal, we can sometimes just give up on the goal rendering what was a problem no longer a problem. Barring that, however, we are stuck with unaddressed problems, and unaddressed problems have a habit of reappearing when ignored. This also applies to problems that are mistakenly thought to be addressed. This is all to say that there is feedback from the world in relation to our problems.

3.4.2. Successful Conceptual Engineering

To see what connection problems and feedback have with externalism, we ought to look at some examples of what are arguably partial successes in conceptual engineering: "gender" and "racism" (and their contemporary conceptual counterparts gender and racism). Both terms have only in the last century come under their current usage. "Gender" was appropriated from its linguistic usage to distinguish between the social and biological differences between men and women by John Money, Joan Hampson, and John Hampson (Meyerowitz 2008, 1354). Feminists found this distinction to be useful and it was picked up in the 70's, with Gayle Rubin being one of the early prominent adopters (Wieringa 1998, 13-15). Money engaged in conceptual engineering by appropriating "gender" from grammar to solve a problem. Money formed the concept gender to pick out the social phenomena gender. He wanted to be able to talk about the social aspects of being a man or a woman without invoking the biological aspects.

Robert Stoller further refined a concept of gender, separating *gender identity* and *gender role* (Bollough 2003, 232). This distinction allowed him to explain how someone could be trans by framing it as a "mismatch" between sex and gender. While there are surely plenty of issues

with this view, one can see how that bit of conceptual engineering has allowed Stoller and those who followed him in that distinction to conceptualize trans people, and more importantly, has allowed many trans people to make sense out of themselves.

Feminist scholars used the distinction to help break biological essentialism; the distinction allowed them to think about socially contingent facts that are supposed to be broadly characteristic of women and reveal them as socially contingent as opposed to the result of biology. The distinction, while certainly challenged in some circles, is relatively commonplace. How this distinction went from academic circles to being part of the parlance of our times can be explained by the way it solved at least these two issues among possible others.

The distinction also helps explain why conceptions of men and women vary among cultures. It also allows one to recognize that insofar as some social aspect of being a man or a woman creates a harm or injustice, that there are ways in which those social aspects can be changed to address such harms or injustices. This use of "gender" implied a binary concept; either you are socialized as a man or woman. A competing concept has gained a foothold in the last decade or so that is decidedly more spectral: this use of "gender" allows one to identify as things outside of man or woman. People identify as androgynous "genders" and "genders" of lower intensity i.e. where they identify less with, and are less guided by, their "gender" relative to the average man or woman.

The first recorded use of "racism" was by Richard Pratt, where he seems to mean something parallel with classism with respect to race in 1902 (Zack 2018, 149). The French counterpart "racisme" was first used to self-describe as part of various political ideologies (Balibar 2008, 1633). During WWII and into the 60's, "racism" was picked up and popularized by academics and social activists as a way of denoting the attitudes and beliefs that belied the

acts of discrimination that were based on notions of race i.e. our contemporary concept of *racism* (Fredrickson 2015, 165). This remains the dominant notion of today.

However, this notion was already being implicitly challenged in the mid 1960's by the man known then as Stokely Carmichael, Kwame Ture in 1966 (Dutch 2019, 306). Ture made a distinction between individual and institutional racism. Individual racism is like the common definition where beliefs and attitudes lead to racist acts. Institutional racism happens within and across institutions and organizations. An individual racist might deny a black person a job, whereas institutional racism could cause hundreds or thousands of black people to go unhired by failing to provide the things needed to get a job in the first place. This distinction implicitly challenges racism as purely a matter of attitudes and beliefs, because institutional contexts may result in racist outcomes with nary a racist attitude or belief. This more inclusive notion of racism, as well as other alternative notions, has become common in at least the English-speaking world, prevalent enough to make changes in the dictionary (Mitchum 2020).

3.4.3. Cappelen's Objections: The Epistemic Objection

The creation of these concepts, and the concepts that built upon them, were not framed as conceptual engineering projects, obviously. Nonetheless, they were self-conscious attempts to change the language to suit human purposes, and they caught on. While it is possible that these were flukes, the better explanation is that these changes let people think about, and let people do things, that they could not think about or do before. There seems to often be a recognition that a newly created concept is useful. Think about any time you've learned a new concept or two. The ways possibilities open up, or things that are fuzzy become clear, invoke a kind of gratification. When we let such concepts guide us, and find the consequence of that guidance somehow satisfactory, we tend to return to these concepts, satisfaction just being any kind of feeling that

reinforces or is correlated with the reinforcement of these concepts' usage. Before we address Cappelen, it should be said that the burden is on anyone who argues against our general ability to conceptually engineer, because my two examples show that we can and do. These are not even necessarily the strongest examples. The cases of *rape* and *marriage*[6] are perhaps even better examples. There are also likely tons of examples across many disciplines, scientific or otherwise. The skeptic has to at least explain the source of the illusion. If it seems like we can purposely alter our concepts, then they had better show why it is merely a seeming and where it comes from. Cappelen never addresses this.

With this laid out, we can address Cappelen more directly. Cappelen starts off by setting out his externalist assumptions:

> (i) Putnam/Kripke/Burge/Williamson-style externalism is part of the metasemantic story: the external environment that speakers are in partly determines extensions and intensions. The relevant elements of the external environment include experts in the community, the history of use going back to the introduction of a term, complex patterns of use over time, and what the world happens to be like (independently of what the speakers believe the world is like). Intensions and extensions don't supervene only on the mental states of the speakers.
>
> (ii) The possibility of massive, fundamental mistakes and confusions about semantic values: most or even all speakers of the language can believe that a predicate F applies to an object, o, but be wrong. They can all want o to be in the extension of F, but wanting o to be F doesn't make it so. They can all be disposed to apply F to o even though o isn't F. Humpty Dumpty was wrong: believing and wanting words to mean something doesn't make it so. (I take it to be a corollary that speakers can all think that F figures in certain explanations or in certain inferences and be wrong.) (2018, 63)

From (i) and (ii), follow Inscrutability and Lack of Control, explicated below, which correspond to the epistemic point and metaphysical point respectively:

---

[6] Arguably $rape_1$ was replaced by $rape_2$ and $marriage_1$ was replaced by $marriage_2$. In the first case, non-consensual sex is still rape even when married and in the second, marriage can apply to both legally binding relations to two people regardless of sex and gender.

> [Inscrutability] An epistemic point: In most cases, the detailed mechanisms that underpin particular instances of conceptual engineering are too complex, messy, nonsystematic, amorphous, and unstable for us to fully grasp or understand.
>
> [Lack of Control] A metaphysical point: The process of conceptual engineering is governed by factors that are not within our control, and no individual or group has significant control over how meaning change happens. Even if we could overcome our epistemic limitations and know all about the relevant factors for a particular case, what we would have knowledge of would be something we had little control over. (2018, 72)

How exactly he moves from (i) and (ii) to the two above points is less than explicit. There are some clues, however. I will attempt to reconstruct the general argument in explicit form, and then I am going to try and show why I think this is a fair reconstruction of his argument. One thing to keep in mind while examining this argument is that, for Cappelen, conceptual engineering is a matter of changing the intensions and extensions of words, and we have no idea how intentions and extensions of words are changed. All we can say, according to Cappelen, is that such changes happen rather frequently. The reconstructed argument is as follows:

1. Reference shifts under externalism are mysterious.
2. If reference shifts under externalism are mysterious, then the best explanation for that fact is that reference shifts are very complex and messy.
3. If reference shifts are very complex and messy, then there is no recipe for causing a reference shift.
4. If we have no recipe for reference shift, then conceptual engineering is inscrutable.
5. If conceptual engineering is inscrutable, then we cannot control reference change.

∴ Conceptual engineering is inscrutable.

I am going to quote and explicate a few passages from the book. The following correlates with premise 1.

> …That understanding the exact underlying mechanism(s) that trigger reference shift is hard is itself an important data point for my theory. The experts on reference don't know how it happens; much less ordinary speakers. (Cappelen 2017, 66)

He goes on to list some of the suggestions that some of the experts have given, but ends by saying that even if one of them or some combination is correct that "...it provides precious little guidance in particular cases." In sum, since the experts can at best give us vague outlines of how reference works, we can infer that the particular mechanics that determine meaning are mysterious. He goes on to give a long quote from Dorr and Hawthorne as an example:

> [T]his argument is motivated by the thought that the number of possible meanings for that expression is enormous. In almost all cases, an expression's actual meaning is surrounded by a vast cloud of slight variants that seem just as well qualified to be possible meanings. And this abundance makes it hard to resist the conclusion that each proposition attributing any one of these meanings to the expression is modally plastic to a high degree. For if there are no differences between the possible meanings that could plausibly explain why a few of them would be much easier to latch onto than the rest, the selection of any one of them as an expression's actual meaning has to depend very sensitively on the exact values of whatever microphysical parameters are relevant to the determination of meaning. Similarly, when there are many slightly different possible contents for a speech act on some particular occasion, and no differences between these contents that could plausibly single out a few of them as being much more apt for being communicated than the rest, the fact that a certain one was singled out as the actual content of the speech act performed on that occasion must depend sensitively on the exact details of the relevant microphysical facts. Let us call arguments of this general kind arguments from abundance.
> (Dorr and Hawthorne 2014: 282)

According to Cappelen, a consequence of this view is that changes in the supervenience base can create change in the semantic superstructure, and that such changes are frequent and can be more or less quick. Still, given that we don't have the ability to track these "microphysical parameters" and don't understand how exactly these parameters determine meaning, we cannot extract a recipe for enacting this change ourselves. I take microphysical facts to include quantum, atomic, molecular, and chemical facts. If these are the kind of facts that Cappelen thinks might explain the semantic facts, then we can see why he believes it to be a messy affair. We can infer that he thinks the lack of a mechanical theory of reference shift can be explained by such messiness i.e. 2. The best support for number three comes from where he quotes Williamson at length:

> Here I find something Williamson says helpful: "Although meaning may supervene on use, there is no algorithm for calculating the former from the latter" (Williamson 1994: 206). In defending epistemicism about vagueness, Williamson also says: "Every known recipe for extracting meaning from use breaks down even in cases to which vagueness is irrelevant. The inability of the epistemic view of vagueness to provide a successful recipe is an inability it shares with all its rivals. Nor is there any reason to suppose that such a recipe must exist" (Williamson 1994: 207). I think this is exactly right and relevant in this context: if we're looking for an algorithm for how to change meanings, we are in effect asking for a recipe for extracting meaning from use, and we have no good reason to think such a recipe exists. (2018, 67)

Cappelen thinks this is true, regardless of whether vagueness holds or not. As this passage clearly states, not only does he think that reference shift is epistemically difficult, it is such a disorderly process that it resists any description more determinate than the vaguest of outlines. For Cappelen, reference change is eternally a black box. 4 follows from 3 definitionally.

> To figure out the current intension of a term, one needs information about past events, introductory events, and communicative chains. Unfortunately, this information is not available and will never be. To effectively change the extension and intension of a term, one must understand the mechanisms of reference change. (2918, 73-74)

This supports 5. For Cappelen, a consequence of inscrutability is lack of control. There might be something like one more premise here. Cappelen seems to think that if reference change could be understood mechanically, then the experts would understand it now. But they don’t. Having defended my reconstruction, there are a number of things to say in response to this argument. First, I concede that the mechanics of how expressions change meaning, and how new concepts spread, are messy, complex processes. That would explain the mystery of reference. Consequently, I will grant premises 1 and 2. 4 is definitional. That leaves premises 3 and 5. Let’s start with 3. Cappelen doesn’t seem to have a good argument for 3:

> First, let's focus on the epistemic point: take one of the terms that are often discussed in connection with conceptual engineering, such as 'person' or 'marriage'. There are two things that we are ignorant of in this context. First, in order to determine the current intension of a term, we would need information

> about past events and communicative chains. This information is not available to us, and we will never have access to it. Second, in order to effectively make a change to the extension and intension of a term, we would need to understand the mechanisms of reference change. However, these mechanisms are unknown to us and may be unknowable. If meaning is based on extremely complex usage patterns over long periods of time and there is no algorithm for extracting meaning from those patterns, it is an illusion to think that we can effectively predict and implement changes. (2018, 73-74)

By Cappelen's own admission, we do not really understand how reference shift happens. If we do not understand the mechanics of reference shift, and the connection between the world and meaning in any substantive detail, then we have no good reason to think the requirements he sets out are the right requirements. Cappelen claims we require something along the lines of introductory events and communicative chains to determine intension. This seems to be asking for too much if we are for precision and specificity. We might only need the broad contours of an expression's history to determine its intension. We can say a lot about the evolution of a species without knowing the exact sequence of events. Maybe we can learn a lot about the nature of a reference change by running very powerful computer simulations. We inquire into a lot of very messy, complex phenomena, but we basically never just give up. We have to remember: externalism as we know it is not that old. The Twin Earth thought experiment was only put forward about 50 years ago. Philosophy of language and linguistics are pretty young. Darwin revolutionized biology. There is always a possibility that some new insight or new framework can make sense of where there was only confusion and mystery before. Cappelen claims the quest is dead without knowing enough about the quest. Even if Cappelen had a good argument, it would have to be a very, very good argument. If there is an answer, and the phenomena are as complex and messy as Cappelen claims, then the likelihood of finding that answer goes to almost 0. This is the kind of thing that we are unlikely to stumble upon without looking for it. We would need a knockdown argument against the possibility of discovering the mechanics of reference

change. Cappelen, at least, has not given us one. He could try to give some kind of argument from the economics of research, and show that it's so difficult that it is a waste of our time. But by Cappelen's thinking, knowing the mechanics of reference change would give a lot more power over reference. That alone would motivate risking the resources. We might also wonder whether, given the admitted difficulties, our framework is wrong. Even if we hold externalism as true, that does not mean that the reference theory of meaning is. If a theory makes it almost impossible to understand the mechanics of something, that constitutes a good reason to abandon that theory.

Let's take aim at 5. We might object that our ability to do something does not depend on knowing the mechanics. For example, philosophers tend to think of themselves as pretty good thinkers. Yet, I am not sure we are anywhere close to having a solid understanding of the mechanics of thought as species, let alone most individual philosophers. Yet, we can think well and control our thinking even. Given that people have implemented exactly these changes, however slowly or organically, it follows that we do not need to understand the mechanisms to be able to do it. Conceptual engineering, whether externalism is true or not, is a subspecies of social engineering when practically pursued. Minimally, it is a close cousin. We know how to social engineer. Marketing departments alone have pulled it off at least a few times. Marketing departments, or their temporal counterparts, convinced women they needed to start removing the hair from their legs and armpits only within the last one hundred years or so, just to give one example. Yet, surely no one can give a recipe, or can say exactly what the mechanisms involved are. Like with conceptual change, we might not ever know, given how messy and complex social engineering is.

Some may argue that, given our success rate in social engineering, we do not have the ability to social engineer. Social engineering is certainly difficult, and our efforts fail far more often than not, but those are not reasons for thinking that we do not have the ability. Consider someone who can split an arrow with an arrow one out of a hundred times. Despite their infrequency, they still count as having some ability because of the difficulty of the endeavor. I could make a million attempts at splitting an arrow with an arrow and never do it, such is the difficulty. And so it goes with conceptual engineering. Cappelen makes a comparison between giving an argument for a conceptual change and an argument for lowering crime in Boston, but the comparison fails for an important reason. When we give arguments for a social engineering project, there is still all the work of implementing the policy, gathering and reorienting resources, and so on. Everyone can think a policy idea is great, but the question after is always going to be "Who's going to pay for it?" All there is for conceptual engineering are arguments and experimenting with the suggested change. Someone might wonder whether Cappelen would have to be right about the comparison if externalism is true.

Here I will return to rejecting 3 by claiming that we do have at least part of the recipe. Whatever meaning is or how reference shift works, one thing we can say with confidence is something has to change in us. Whatever an expression refers to, and whatever it comes to refer to, the change either comes by chance or by reward. Changes in expression are likely often due to what we might call "arbitrary drift," and other changes come about because they somehow help us. There's some kind of reward or benefit that comes about from the shift in expression, reference or otherwise.

Our expressions influence how we interact with the world and ourselves: the way we think, the way we act, how we feel, and how we frame and relate to the world. These changes

can discourage or encourage the employment of expressions, both intra- and interpersonally. If you think that we are physical beings, whatever else we might be, then changes in expression have to be partially explained in terms of brains and bodies. Even if reference or meaning is purely abstract, there has to be some physical correlate or counterpart that allows our meaty being to change in response. My theory gives us a template of how this happens.

First, we encounter a problem. We employ our concepts, or expressions, or what have you, to try and address the problem. When we invent or discover something that successfully addresses that problem, we tend to stick to that, all else being equal. It is the feedback loop between our concepts and expressions, our changed thought/behavior, and the rest of the world, that at least partially anchors meaning or reference. This has to be part of the story, because only something along these lines can explain how we adapt our concepts, language, and conceptually/linguistically guided behavior to the world.

3.4.4. The Metaphysical Objection

Now we can switch over to his other argument which, according to my charitable reading, is the following:

1. If externalism is true, then we have no control over the reference-fixing facts.
2. Externalism is true.

∴ We have no control of the reference-fixing facts.

I will not deny 2, so I will focus on 1. The first issue is that Cappelen never makes clear exactly what he means by control. Under some conceptions of control, we certainly do not. As I discussed above, the chances of success for any given proposal for conceptual change are not going to be high. Thus, if we mean by control something like being able to reliably succeed, then no, we do not have control. However, that is a high standard for control given the context, and we do not have only two options: control or being at the mercy of the whims of the universe.

There is an entire continuum there. There are degrees of influence. It is hard to say exactly how much influence qualifies as control or helplessness. I think it is more reasonable to think in terms of our ability to induce conceptual change or reference shift over time for the purposes of solving a problem. Like splitting an arrow with an arrow, it might take quite a few tries, but just making it happen counts for something. Most big social and scientific problems work this way, too. We keep trying until we get somewhere. Even if we do not have the kind of control when we move our limbs or institute a change in governmental policy does not mean we do not have control in some robust sense. As many examples can show, we have induced conceptual or linguistic changes to solve problems.

One thing Cappelen does not make clear is whether he means tokens or types of expressions. We could say that "cat" as a type refers to a cat without every token referring to a cat, because a particular token of "cat" can, for example, be used in a figurative way. We might be inclined to think that Cappelen could accept something like this, but Cappelen claims we have not a lick of control even when stipulating or when a court interprets a statute:

> It's then natural to think that this will determine the meaning of 'intuition' at least as it occurs in my book... Isn't that a little bit of control? On the view proposed here, the answer is no. All you have achieved is to define a word in a certain way on page 2 of your book. Saying you want a word to mean something doesn't make it so. You have not changed the meaning of the word 'intuition'. What you say when you use that word is governed by what the word means, not by what you want it to mean. The rest of your paper is in English, not in some new language you created on page 2.2 At best your definition will tell charitable readers how to get at the speaker's meaning, i.e., what you had in mind. However, even that isn't a move into a safe space in which we have control. The content of what's called 'speaker's meaning' is just as externalistically determined as linguistic meaning: we have no more control over that content than we have over what we say when we utter sentences in a public language (2018, 75-76)

That suggests that he thinks the referent of even every token is fixed, given a certain place and time. This would make figurative language, or at least some cases of figurative language, entirely perplexing. It might be relatively easy to explain well-worn pieces of

figurative language as some kind of reference expansion, something that allows reference shift under special contexts. It seems like it is much harder to say what's happening when figurative language is novel. While figurative language can be unclear, especially on first impression, we often understand it instantly, even when novel. Insofar as novel figurative language is used to communicate anything moderately specific, it certainly seems like there is some kind of contextual reference shift that is happening intentionally.

Cappelen makes no effort to explain this away. Cappelen seems committed to a particular understanding of what externalism entails, even though he claims we cannot say much about how it works. I call this version of externalism "stubborn externalism," because reference shifts happen on the world's terms and no one else's. We have no reason to think, however, that if externalism is correct, that it is a particularly stubborn version of it. It seems like insofar as Cappelen's claims about our epistemic position with regards to the metaphysics of meaning and reference are right, we could not possibly be in a position to know how stubborn the right version of externalism is. We also don't know what kind of facts meaning is anchored in: facts about natural kinds, physical facts, biological facts, social facts, or even metaphysical facts. We do not know if all kinds of facts are equally resistant to human will. Cappelen seems to have "water is H2O" as his exemplary case. Even if some words or concepts are grounded in natural or physical facts, that should not suggest all are. There's plenty of reason to think that water's reference is fixed in a way that other expressions or concepts are not. Historically, the correlation between the stuff we call "water" and H2O has been very strong. It also a specific substance that our bodies need; it is vital.

No substitutes will do. The properties of H_2O give oceans, lakes, streams, rain, and other forms of what we call "water" the properties that cause us to classify all those things as "water."

It is H_2O that makes our biological chemistry possible. Compare this to something like jade. It turns out that "jade" actually refers to two different minerals, nephrite and jadeite. At least for a layperson, there is no need to distinguish between the two, because they bear sufficiently similar properties with respect to their aesthetics. They look similar enough to keep the same name. The discovery of jade being two minerals did not change usage, because the usage of "jade" works well for most people's purposes. We could imagine "water" becoming jadified in a Twin Earth scenario, where they start importing XYZ to our Earth for its supposed healing properties, being marketed as "Earth 2 Water." Of course, it would have no healing properties, since XYZ has identical macroproperties to H_2O, but having identical macroproperties is exactly what makes such a reference shift likely to take place. "Water" would become like "jade" in terms of usage at least, because calling them both "water" only ever helps the layperson. The layperson can drink, swim, bathe, and so on in XYZ as well as H_2O. Part of the reason "water" is anchored in such a fixed way is that XYZ is (probably) physically impossible. It is the special role H_2O plays in our daily lives that fixes "water" so tightly to it. Of course, specialists might have good reason to be specific, to break with wide usage because, for the specialist, there could be differences that make a difference. Even concepts that tightly fixed to physical facts are not all equally resistant to human will; such concepts can be altered or replaced, given the right context and motivation.

Then, there are social and abstract concepts. It is reasonable to hypothesize that a lot of social and abstract concepts and expressions are grounded in social facts, the kind of concepts that have historically varied a lot across cultures. There, only social facts would need to be changed, and all that it takes to change such facts is that a community be sufficiently unified and sufficiently motivated to make a change — generally, because such a change addresses some problem. I say "all that it takes," as if that sufficient unification and motivation is not some huge

demand. It is a huge demand, but one that is well within realm of metaphysical possibility. All sorts of things might prevent such unity and motivation, but there is history of societies making such purposeful changes. People can consciously choose to start employing a new word or concept, and others might pick it up unconsciously. They will only be motivated if there is some problem they have to address where changing the reference will address that problem. In adopting some change, the ongoing interaction between the world, including other people, that the change creates can reinforce or degrade the continued usage of that concept.

One subtle assumption that Cappelen might be making is that semantic facts are anchored in a determining way, as opposed to a constraining way. He seems to think that the facts that anchor meaning and reference make an expression mean or refer to one thing and one thing only, at least in a given instance. They determine the reference of an expression. However, it might just as well be that they merely constrain meaning and reference. That is to say, there is a range or set of possible meanings or references for an expression. This might sound like Dorr and Hawthorne from the quote above. However, what should be taken as the actual content of some expression is one that is determined by the microphysical facts. What I am proposing is that there is not matter of fact of what the actual content is, only a matter of fact about which ones are not the content. There are things that an expression cannot mean, but a number of things an expression could mean or reference. This could be what makes language in one way seem so flexible and rigid at the same time. We cannot choose to make an expression whatever we want because use is bounded by external anchors, but within those bounds we can manipulate the meaning and reference of expression in a number of different ways. A constraining externalism would likely be a lot more amenable to conceptual engineering.

An assumption that Cappelen arguably makes is that such change of words or concepts would have to involve change in meaning or reference, even at the beginning of some wider adoption. However, one way in which externalism might hold while allowing for these kinds of changes would be that initially meaning or reference does not change. Something else is happening, something akin to stipulation at a larger scale. For example, it is or is something like a convention-making expression: that before whatever has to happen to create the change in meaning in reference, people speak or write words to say "Yay to using this sound to invoke this idea or concept or image" and insofar as the right things are invoked such that problem at hand is addressed, that expression spreads until it starts to settle and there is finally a shift in meaning.

3.5.0. Addressing the Objection from Discontinuity

Addressing changes in topic can also be accomplished by appealing to the cognitive role theory. First, we have to think about what a topic is. If we are on the topic of biology, and I start talking about the geophysics of Jupiter's moon Io without somehow tying that back into how life works or what the extension of life is, then I have changed the topic or attempted to do so[7]. Insofar as the topic of biology is bound to the extension and operations of life, then anything that is not explicitly in the bounds of how life works or life's extension is off-topic. A topic just sets the bounds of what is relevant, what is assertable in some context. It is the linguistic counterpart to a fence.

In casual conversation, there isn't usually even a need for topics unless some kind of disagreement arises (which might render the conversation non-casual). A topic is needed in disagreement, because any point brought up needs to be sufficiently relevant to have any hope of resolving the disagreement, if it can be resolved at all. Even here, though, an accusation of being

[7] Obviously, my interlocutor after waiting a while to see if I tie it back into the topic at hand might say something to the effect of "That was interesting but let's get back on topic."

off-topic can itself be challenged. Even if I bring up Ionian geophysics when the topic is biology, I might try explaining how there might be some kind of exotic lava life. The accuser might draw the line at cellular life. The so-called winner of this clash will be partly, if not primarily, determined by the plausibility of lava life and the strength of similarity between Earth life and non-cellular Ionian life, but that such negotiations are legitimate shows that the bounds of the topic are not fixed. A disagreement is just a certain kind of problem. Topics are important for cases that go beyond mere disagreement; that is cases where there is a problem that exists beyond disagreement. There might be disagreement with respect to the problem, but even absent disagreement there is a problem. Again, topic marks the bounds of relevance. In this case, it marks what could be relevant to addressing the problem. Whether life should extend to non-cellular life is going to depend on the context. If we are just talking about what we can expect out of life on Earth, then life being bound to just cellular life might be best. If the context is astro- or xenobiology, what we can expect from the possible entities across the universe, then we might be unwisely excluding all the non-cellular entities that are otherwise sufficiently similar to cellular life.

Kraus s's implied proposal to restrict the use of "nothing" to the lack of stable material objects fails i.e. restricted nothing. For those who are interested in the question "why is there something rather than nothing," the desire is to understand how ANYTHING can be, stable or unstable, material or otherwise. Krauss could put forward an argument showing that the existence of fields and particles are brute, and that any questions about their origins are either meaningless or entirely uninteresting. An argument like that could give us a reason to switch restricted nothing: it would show us that such questions cannot give us what we want, that they cannot lead us to solving any problem. Such an argument would show us that there is no

problem; that the restricted nothing is a better concept. Topics, too, can be replaced, if there is some other topic that better serves that cognitive role. If we replace truth with *ascending* and *descending truth* to fix the family of paradoxes, then there are two possibilities: either the topic has not changed because the topic was "What concept can we employ that serves the alethic role without paradox?" or we have replaced the old topic with that topic. We can replace topics themselves because topics themselves are cognitive tools in service of a cognitive role.

3.6.0. Addressing the Objections from Loss

As it turns out, three of the objections turn out to be variations of what we might call an objection from loss, despite one of the objections being against feasibility and the other two being against recommendability. Each objection is an articulation of a worry about losing something important, whether that be belief, knowledge, or whole swaths of our conceptual apparatus. In the case of the objection from fundamentality, the loss is presented as simply not possible, whereas in the others, the loss is presented as possible, but impermissible.

If anything at all is fundamental, it is the cognitive roles that *truth* plays, not *truth* itself. There is little reason to think that whatever the exact concept that "truth" refers to is the same exact concept that other cultures and languages refer to by their respective counterparts to "truth" insofar as they have such counterparts. This goes both across cultures and languages, and along the histories of individual languages and cultures. After all, even analytic philosophy cannot agree on the nature of the concept even now. Both the logic of *truth* and the philosophy of *truth* are very contentious. We can expect things to differ to a greater extent across groups who share far fewer of the same assumptions.

There is reason to think that different cultures and languages have counterparts to *truth* and "truth." Although not the exact same concepts as *truth*, they still play alethic roles. These

concepts are playing alethic roles, certainly what I called the (Conciliation) role in 3.3.0 and probably something akin to the (End of Inquiry)[8] role. These roles might be fundamental. If they are, we are allowed to replace truth if the replacement can play such those roles. What this suggests is that human psychology is not so fine grained as to have very specific concepts built in. If anything is built in, it's slots that concepts need to fill: a doxastic slot, an alethic slot, etc. Whatever else human psychology might be, it is plastic.

So what is important are not the tools per se. What's important are the roles that the tools play, as well as having good tools to fill those roles. Losing *knowledge*, *belief*, or any other number of concepts ought to be considered a loss, under the conditions that there are no other tools that can play the roles that they play as well as they do. These concepts do important theoretical and practical work for us. Under this light, the objection from centrality also loses its force, because any concept constituted by truth can itself be replaced by a different concept (knowledge-2, belief-2, etc.) and so down the chain. These new concepts may functionally be identical except perhaps in very specific circumstances, if even then. With knowledge-2, we study epistemology-2. If a change did create significantly different functioning in a bad way, that would count as reasons to refuse such a change. We have no reason, for example, to think a conception of knowledge constituted by *ascending* and *descending truth* would function any worse than one constituted by truth, except that it could perhaps avoid certain paradoxes, even if by the conceptual identity principle, it is technically a different concept. Ultimately, we should worry no more about replacing these concepts with either improved versions of these concepts or wholly new concepts than we worry about replacing one of our claw hammers with a better claw hammer or even a nail gun.

[8] To clarify, (End of Inquiry) is a call back to the platitude so I am using the platitude to name the role vs. "the conciliatory role" which is a role I named.

3.7.0. Addressing the Objection from Underjustification

The objection from underjustification amounts to saying that we lack the justification to replace truth. More specifically, only inconsistent beliefs created by truth's inconsistency would be enough to motivate replacing truth with respect to its inconsistency, and inconsistent concepts likely do not create inconsistent beliefs. Belief is an unlikely candidate for what inconsistent concepts amount to, because then competence with inconsistent concepts would entail that we are irrational, since we would hold contradictory beliefs. They are much more likely to create inconsistent entitlements, rules, or dispositions, since they can conflict without us being irrational per se.

There are a number of things to say here. First, there is reason to doubt that only contradictory beliefs could give us enough justification to replace truth. It is, of course, orthodoxy that contradictory beliefs are the mark of irrationality. We can concede as much while not denying the justifying force of contradictory entitlements, rules, or dispositions. Whatever other tasks philosophy might involve, the notion that philosophy involves keeping our theoretical and conceptual house in order is a pretty plausible one, and it is never a matter of simply keeping your beliefs from contradicting one another. It is a multifaceted affair: making good inferences, working out sub-contradictory tensions, gaining the right sensibilities of inquiry, and keeping our actions and beliefs lined up. Trying to keep entitlements, rules, dispositions or what have you orderly seems natural to rationality. Plus, there is no obvious threshold as to when something gets replaced, just like there is no obvious point where an established scientific theory should be replaced by a newer, more powerful theory. We, individually and collectively, have to make that call in those particular cases. It cannot be said a priori at what point logicians or linguists should be motivated to replace truth, if they should ever do so. Under a framework like mine, we don't

need to worry about justification in this way. Concepts are tools, and we change or replace our tools when we see some benefit. As long as there really is a benefit, and that benefit is such that it motivates people to take up the new or modified concept, there is nothing else we can ask for. We can trust people, and experts in particular, not to work any harder than they when it comes to adopting conceptual change.

Whether the objection ultimately succeeds or not really comes down to whether truth's inconsistency turns out to be a genuine problem. Problems, as I mentioned, do not have be recognized as problems to count as problems. Part of what makes a problem a problem is that it cannot be trivially addressed. Not just anything will address it, nor will ignoring it address it. It demands to be dealt with in a substantive way. Is *truth*'s inconsistency a problem in the relevant sense? It cannot be definitively said, but people have grappled with paradoxes surrounding *truth* for a long while now. They continue to haunt us, and we continue to discover new paradoxes. Considerable ink has been spilt on these paradoxes. Obviously, it seems unlikely anyone will starve due to the inconsistency of truth, but no one has probably starved because sometimes knowledge is lucky either. These are intellectual problems, philosophical problems. This is the world resisting our efforts to intellectually systemize it. Starvation, cancer, and oppression do not have a monopoly on problems. If such things did, almost no philosophical problems would count as problems. The fact that the inconsistency confronts different people across different times suggests that it is a genuine problem. Maybe it will turn out to be just a fad. Maybe we will simply choose to ignore it and be no worse off. Maybe the cure is ultimately worse than disease. We cannot say, but as long as it presents itself as a problem to us and replacement exists as a means of addressing it, then we are justified in attempting replacement.

3.8.0. Conclusion

Here, we will finally address the feasibility and recommendability of replacing *truth*, and say something about what it means for replacement as a general strategy. Under the cognitive toolkit framework (CTF), conceptual engineering can be thought of as a special case of cognitive engineering, and concepts (or their mental counterparts) as a species of cognitive tools. Cognitive tools fulfill cognitive roles in the service of addressing some problem, and hence, the feasibility and recommendability of the replacement strategy follows almost trivially from CTF. Under CTF, all concepts are cognitive tools that can be replaced in any cognitive role that they play, provided there are tools that better serve that role. Replacing *truth* as strategy is theoretically feasible, because cognitive tools are replaceable by nature. The cognitive tools we use now are not the ones we used before, and our own current cognitive tools will all likely be replaced in time. In CTF, there are no strong in-principle reasons why a cognitive tool cannot be replaced. The objections that would amount to strong in-principle reasons either fall short or fail altogether. The strongest objection, the objection from externalism, is at best insufficiently general. Although there may be cases where external anchors determine the problem and constrain the cognitive role in such a way that only our current concept can fulfill the role, with CTF, that would be the exception. To the extent that external anchors generate limitations, those limitations will be built into the cognitive role and its respective problem. Concepts that are not sensitive to those anchors simply fail to fulfill the role and hence, fail to address the problem. The objection from fundamentality fails because cognitive tools are not fundamental. At best, some cognitive roles might be. The discontinuity objection fails because topics are just cognitive tools. They are not the ultimate arbiters of what concepts are acceptable.

While I made it explicit that practical feasibility was not the focus of the thesis, it is still worth saying how it interacts with CTF. Practical feasibility could be said to be more demanding.

Concepts that are too complex, too vague, or insufficiently distinct from the target concept might interact with the relevant communities in such a way that they fail to catch on, even if they are highly effective at fulfilling the relevant role in the hands of an expert. Proposals must be communicated. Any proposals for conceptual change must be introduced to others. More effective communication, both in terms of reach and quality, will affect how likely it is that the relevant communities will adopt a proposal, but the basic requirement is that the conceptual engineer shares the proposal with at least some of the relevant communities.

Replacing *truth*, then, is a feasible strategy, because *truth* is a cognitive tool serving the alethic role and can be replaced by any other cognitive tool. In fact, anything that can address the alethic paradoxes will be a distinct cognitive tool and will count as a replacement. Therefore, all the proposals are something different from *truth* as we know it. Even a paraconsistent resolution would mean replacing *truth,* because the term "true," used in a paraconsistent way, is compatible with the described sentence being "false." *Paraconsistent truth* is not the same as *truth*. Hence, everything I have written about cognitive tools and concepts under CTF applies to all means of addressing the paradoxes.

Replacing concepts, including *truth*, is almost always going to be recommendable, as long as there are genuine competitors to the received concept. Again, recommendability does not mean that one should replace a concept, only that replacement has a prima facie deliberative considerability. One would not be wrong to recommend the change, apart from whether such a change is actually made. Granted, there might not necessarily be any competitors. The cognitive role might be satisfactorily fulfilled. The problem, as we know it, might be satisfactorily addressed. This is not the case with *truth*, and work continues to be done on addressing the paradoxes, which naturally generates many contenders. Paradoxes are not the only reason to

replace *truth*. All the different theories of *truth* count as contenders for replacement insofar as we treat the correspondence theory as the received concept of *truth*, each of them proposing a different concept we ought to invoke with the term "truth."

Under CTF, we can conserve roles while changing cognitive tools, ensuring that their functions are preserved while allowing for conceptual improvements. The objection from centrality fails because we can maintain all of the roles, and so changing *truth* as the fundamental concept to ascending and descending *truth* will not change how the new concepts, like ascending-and-descending-belief, function relative to the belief concept we are familiar with. The same can be said about epistemic loss. The mental entities that are picked out by knowledge still fall under the epistemic role that knowledge served i.e. the knowledge tokens will count as ascending-and-descending-knowledge tokens. The objection from underjustification fails because, insofar as *truth*'s inconsistency is seen as a problem by the relevant communities and replacement is the go-to strategy, the effort is justified. Perhaps, it will turn out not to be the problem we thought it was, but the mere possibility that we are wrong about it being a problem does not undermine the justification of the effort.

What does *truth* tell us about replacing other concepts? Under CTF, there is no reason to think that *truth* is some special case. If we fail to distinguish the concepts and the roles that they play, we can see why *truth* might give us special worry, but under CTF, *truth* is not unique. Is there any reason to think any class of concept or some particular concept will be irreplaceable? Yes, but only in the sense that some of the concepts we have are simply the best that we can do. *Water-as-H_2O* on a twinless Earth in a universe where no other compound can replicate the exact properties of H_2O is probably the only viable conception of water we can have. We can no more replace that concept than we can climb a mountain higher than its peak. So, only something like

peak concepts are irreplaceable, but we cannot tell beforehand which concepts are peak concepts. False positives are always possible. Our search for better concepts requires both vigilance and experimentation, just as many of our projects do.

What I have attempted to lay out here is a plausible forward for conceptual replacement and conceptual engineering by providing a structured but open approach to evaluating replacement in general, as well as in particular cases. That said, there are a number of other areas that could be more deeply explored. One area worth exploring would be the relationship between my framework and the different conceptions of concept, what implications those particular conceptions have for CTF and conceptual engineering Another rich area for research would be a project providing a more exhaustive account of problems and the precise relationship between problems, cognitive roles, and cognitive tools. That would pair well with a more expansive application of the framework to a series of case studies, looking closely at the dynamics of competing cognitive tools, cases where cognitive roles themselves are replaced, and how, in detail, cognitive tools and roles interact with problems and how that drives conceptual change.

REFERENCES

Balibar, Étienne. 2008. "Racism Revisited: Sources, Relevance, and Aporias of a Modern Concept." PMLA, vol. 123, no. 5, pp. 1630-1639.

Burgess, A., and D. Plunkett. 2013. "Conceptual Ethics I." Philosophy Compass, vol. 8, pp. 1091-1101. https://doi-org.ezproxy2.library.colostate.edu/10.1111/phc3.12086.

Bullough, Vern L. 2003. "The Contributions of John Money: A Personal View." Journal of Sex Research, vol. 40, no. 3, pp. 230-236.

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

Carnap, Rudolf. 2003. The Logical Structure of the World: And, Pseudoproblems in Philosophy. Open Court Publishing.

Dorr, Cian, and John Hawthorne. 2014. "Semantic Plasticity and Speech Reports." The Philosophical Review, vol. 123, no. 3, pp. 281–338.

Eklund, Matti. 2014. "Replacing Truth." Metasemantics: New Essays on the Foundations of Meaning, edited by Alexis Burgess and Brett Sherman, Oxford University Press, pp. 293-310.

Fredrickson, George M. 2015. Racism: A Short History. Princeton University Press.

Greenough, Patrick. 2019. "Conceptual Marxism and Truth: Inquiry Symposium on Kevin Scharp's Replacing Truth." Inquiry, vol. 62, no. 4, pp. 403-421.

Gupta, Anil, and Nuel Belnap. 1993. The Revision Theory of Truth. MIT Press.

Haslanger, Sally. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" Noûs, vol. 34, no. 1, pp. 31–55.

Huizinga, Johan. 1955. *Homo Ludens: A Study of the Play-Element in Culture*. Translated by Cécile Seresia. Boston: Beacon Press.

Hatch, Justin D. 2019. "Dissociating Power and Racism: Stokely Carmichael at Berkeley." Advances in the History of Rhetoric, vol. 22, no. 3, pp. 303-325.

Krauss, Lawrence M. 2012. *A Universe from Nothing: Why There Is Something Rather than Nothing*. Free Press.

Lynch, Michael. 2009. Truth as One and Many. Oxford University Press.

Meyerowitz, Joanne. 2008. "A History of 'Gender'." The American Historical Review, vol. 113, no. 5, pp. 1346-1356.

Mitchum, Kennedy. "Racism Definition: Merriam-Webster to Make Update After Request." 2020. BBC News, 10 June 2020, www.bbc.com/news/world-us-canada-52993306.

Priest, Graham, Koji Tanaka, and Zach Weber. 2025. "Paraconsistent Logic." The Stanford Encyclopedia of Philosophy (Spring 2025 Edition), edited by Edward N. Zalta & Uri Nodelman, forthcoming. https://plato.stanford.edu/archives/spr2025/entries/logic-paraconsistent/.

Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy & Public Affairs* 13 (2): 134–171.

Ramsey, William. 2022 "Eliminative Materialism", The Stanford Encyclopedia of Philosophy (Spring 2022 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2022/entries/materialism-eliminative/>.

Scharp, Kevin. Replacing Truth. 2013. Oxford University Press.

Scharp, Kevin. "Conceptual Engineering for Truth: Aletheic Properties and New Aletheic Concepts." 2020. Synthese, vol. 198, no. Suppl 2, pp. 647-688.

Schwitzgebel, Eric. 2023. "Belief." The Stanford Encyclopedia of Philosophy (Fall 2023 Edition), edited by Edward N. Zalta & Uri Nodelman. https://plato.stanford.edu/archives/fall2023/entries/belief/.

Simion, Mona. 2018. "The 'Should' in Conceptual Engineering." Inquiry, vol. 61, no. 8, pp. 914-928.

Simion, Mona, and Christoph Kelp. 2020. "Conceptual Innovation, Function First." Noûs, vol. 54, no. 4, pp. 985–1002.

Strawson, P. F. 1963. "Carnap's Views on Constructed Systems versus Natural Languages in Analytic Philosophy." The Philosophy of Rudolf Carnap, edited by Paul Arthur Schilpp, 1st ed., Open Court, pp. 503-518.

Tarski, Alfred. 1956. "The Concept of Truth in Formalized Languages." Logic, Semantics, Metamathematics: Papers from 1923 to 1938, Oxford University Press.

Wieringa, Saskia E. 1998. "Rethinking Gender Planning: A Critical Discussion of the Use of the Concept of Gender." Gender, Technology and Development, vol. 2, no. 3, pp. 349-371.

Williamson, Timothy. 1994. Vagueness. Routledge.