Scharp on inconsistent concepts and their engineered replacements, or: Can we mend these broken things?

Scharp on inconsistent concepts and their engineered replacements, or: Can we mend these broken things?

Mark Pinder

Mark.Jonathan.Pinder@gmail.com

The Open University

The Version of Record of this manuscript has been published and is available in INQUIRY via

http://www.tandfonline.com/10.1080/0020174X.2019.1688181

Abstract. Kevin Scharp's influential work on the alethic paradoxes combines an extensively developed inconsistency theory with a substantial conceptual engineering project. I argue that Scharp's inconsistency theory is in tension with his conceptual engineering project: the inconsistency theory includes an account of concepts that implies that the conceptual engineering project will fail. I recommend that Scharp revises his account of concepts, and show how doing so allows him to resolve the tension. The discussion is important for ongoing work on conceptual engineering. Firstly, it is important to get clear on whether Scharp's conceptual engineering project is—or could be—successful. Secondly, the issues discussed herein may generalise to other conceptual engineering projects, such as explication and ameliorative projects. In particular, the discussion has implications for how conceptual engineers can think about concepts and their relation to how we use words.

Key Words. Scharp. Conceptual engineering. Inconsistency theory. Truth. Concept.

1. Introduction

Consider the following sentence, λ:

> λ is not true.

If λ is true then, given that λ says that λ is not true, it follows that λ is not true. But, if λ is not true then, as that is just what λ says, it follows that λ is true. So λ is true if, and only if, λ is not true. By

considering cases, an explicit contradiction can be derived. This is the liar paradox, one of many paradoxes involving truth and falsity—the so-called *alethic paradoxes*.

Kevin Scharp's influential work on the alethic paradoxes combines an extensively developed inconsistency theory with a substantial conceptual engineering project.[1] An *inconsistency theory*, in the present sense, is any view on which TRUTH and/or related concepts are deemed to be, in some sense, inconsistent. According to Scharp, the alethic paradoxes show that TRUTH is inconsistent and, in consequence, TRUTH is unsuitable for various theoretical roles—such as playing the central role in a semantic theory for natural language. He subsequently engineers a team of replacement concepts—ASCENDING TRUTH and DESCENDING TRUTH—to play such theoretical roles in place of TRUTH.

In the present paper, I argue that Scharp's inconsistency theory is in tension with his conceptual engineering project: the inconsistency theory includes an account of concepts that implies that the conceptual engineering project will fail. I recommend that, to deal with the tension, Scharp revises his account of concepts. I sketch one way of doing this, and show that the resulting account provides Scharp with the resources to resolve the tension.

This discussion is important for ongoing work on conceptual engineering for at least two reasons. Firstly, Scharp's conceptual engineering project is one of the principal examples of conceptual engineering in philosophy, and it is important to get clear on whether it is—or could be—successful. If it is successful, then Scharp's project illustrates this powerful methodology in action; if not, then it illustrates the pitfalls and problems that the methodology can face. Either way, there is much to learn. Secondly, the issues discussed herein may *generalise*. The specific tension that Scharp faces threatens to apply, if less acutely, to other conceptual engineering projects—such as explication and ameliorative projects—and the *kind* of resolution I suggest might be adapted to such projects. More precisely, the discussion will have implications for how conceptual engineers can think about *concepts* and their relation to language.

I begin (§2) by sketching Scharp's inconsistency theory and conceptual engineering project. I then (§3) develop the tension between these two aspects of Scharp's view, before (§4) revising

[1] See Scharp 2013, 2019.

Scharp's account of concepts and (§5) resolving the tension. I close (§6) with some comments about how the discussion generalises.

2. Scharp on alethic paradoxes

2.1 The inconsistency theory

According to Scharp, the alethic paradoxes arise because the concept TRUTH is *inconsistent*.

Here, "inconsistent" is to be understood as a term of art. Scharp understands concepts to be constituted, in part, by principles. To say that a concept is *inconsistent* is to say that its constitutive principles are *jointly inconsistent (with the set of all facts)*. For example, Scharp claims that all instances of the following two unrestricted schemas are constitutive of TRUTH:

(T-In) If ϕ, then $\langle\phi\rangle$ is true.

(T-Out) If $\langle\phi\rangle$ is true, then ϕ.

Here and below, "$\langle\cdot\rangle$" is a name-forming device. Now, consider the biconditional obtained by substituting λ into (T-In) and (T-Out):

(T_λ) $\langle\lambda$ is not true$\rangle$ is true if, and only if, λ is not true.

While (T_λ) per se is not inconsistent, it is jointly inconsistent with the fact that λ is identical to $\langle\lambda$ is not true$\rangle$. It is in this sense that, on Scharp's view, TRUTH is inconsistent.

A difficult question, which any theory along these lines must tackle, is what grounds the claim that a concept, *c*, has constitutive principles, $P_1,\ldots,P_n$. Call this the *grounding question*. It is standard to answer the grounding question by appealing in some way to conceptual competence. Perhaps, say, *c* has constitutive principles $P_1,\ldots,P_n$ in virtue of conceptual competence with *c* being a matter of having the right kind of dispositions towards $P_1,\ldots,P_n$—dispositions-to-accept, dispositions-to-believe, or something else.[2] Or, perhaps, *c* has constitutive principles $P_1,\ldots,P_n$ in virtue of

[2] See, e.g., Eklund 2002a,b, 2007.

conceptual competence with *c* being a matter of standing in the right kind of cognitive relation to $P_1,\ldots,P_n$.[3]

Scharp's considered view is that these approaches to the grounding question are "a mistake" (2019: 445).

> Constitutive principles have little or nothing to do with competence or concept possession. If you were in a conversation where someone asserted 'not all vixens are vixens', you would probably wonder whether you mean the same thing by some of those words, but you would not infer that this person does not possess the concept of vixen. After all, if your interlocutor means something else by 'vixen', then this sentence isn't about vixens. You can't infer anything about the concepts possessed by this person. (p. 444)

Scharp recommends an alternative approach. The intuitive idea is this: what grounds the claim that each instance of (T-In) and (T-Out) is constitutive of TRUTH is that, if someone were to deny an instance thereof, we would likely wonder what they meant by "true". The details are worth spelling out.

To begin, Scharp claims that there are two major modes of conversation.

> The first is regular (or transparent) mode, where the conversation goes along exactly as you would expect a well-behaved conversation to progress. The participants take turns making claims about the topic of conversation […]. But the second mode for a conversation is *meaning reflection*, where one or more participants questions whether everyone means the same thing by one or more of the words or sentences involved in the conversation. (2019: 439)

Utterances of certain sentences can lead conversational participants to initiate this second mode of conversation. For example, if I were to utter:

(1) French red wines tend to be hoppy,

[3] See, e.g., Patterson 2007, 2009.

you might question whether we mean the same thing by "hoppy", initiating meaning reflection with respect to "hoppy". Scharp calls the *negations* of such sentences *constitutive sentences for a word for a person*.[4] That is,

(2) French red wines don't tend to be hoppy

is a constitutive sentence for "hoppy" for you just in case: an utterance of (1) would lead you to initiate meaning reflection with respect to "hoppy".

Note that a sentence can be constitutive for different words for different people. Suppose that, were I to utter (1) to Anya, she would initiate meaning reflection with respect to "wine". Then, on that supposition, (2) is a constitutive sentence for "wine" for Anya. And suppose that, were I to utter (1) to Catalina, she would not initiate meaning reflection at all. Then, on that supposition, (2) is not a constitutive sentence for "hoppy", "wine" or any other word, for Catalina.

Now, Scharp uses the idea of *constitutive sentences for a word for a person* to introduce *constitutive principles for a word for a person*. First, Scharp claims that some constitutive sentences (for a word for a person) are *stronger* than others. The harder it would be for you to accept my utterance of (1) without initiating meaning reflection with respect to "hoppy", the 'more constitutive' (2) is for "hoppy" for you (2019: 439). According to Scharp, there is an important threshold here: if (2) is a sufficiently strong constitutive sentence for "hoppy" for you, then it meets the threshold and counts as a *constitutive principle* for "hoppy" for you; otherwise not.

There are some obvious questions to ask about this threshold. In particular, we might ask where the threshold is (how strong must a constitutive sentence be to be a constitutive principle?) and why it is wherever it is (why is the threshold *there*?). There is an *easy* way to answer these questions: the threshold is in the lowest possible position, so that every constitutive sentence counts as a constitutive principle. This gives the threshold a determinate position and, given that the position is at an extremum, makes that position non-arbitrary. Now, Scharp does not intend to give these answers,

[4] I assume throughout that '*is the negation of*' is a symmetric relation, so that (for example) "grass is green" is the negation of "grass is not green".

which is why he introduces a threshold rather than identifying the constitutive sentences as the constitutive principles at the outset. And, as such, he owes us his view on the matter. But, nonetheless, the existence of easy answers suffices for us to put these questions aside herein: Scharp *can* provide a principled location for the threshold, regardless of whether it is where he wants it to be. For ease of discussion, I henceforth put this issue aside, assuming that: *P* is a constitutive principle for *w* for *x* just in case a denial of *P* would lead *x* to initiate meaning reflection with respect to *w*.

Scharp uses constitutive principles (for a word for a person) to 'define' concepts via an abstraction principle.

> For example, where x and y are lines, the direction of x = the direction of y iff x and y are parallel. This well-worn example features an abstraction principle for defining directions. […]
>
> We can utilize the same sort of principle to define concepts in terms of constitutive principles for words. If w_1 and w_2 are words, then,
>
> (Concept) The concept expressed by w_1 = the concept expressed by w_2 iff the set of constitutive principles for w_1 is identical to the set of constitutive principles for w_2 […].
>
> We treat this as a definition of 'concept'. (2019: 440)

Although Scharp does not make the point explicit, he subsequently takes the constitutive principles for w_1, w_2, to be constitutive principles for the concepts expressed by w_1, w_2, respectively.[5]

Note that (Concept) as it stands does not account for the fact that constitutive principles for words are, on Scharp's view, relativized to people. This relativization can be incorporated unproblematically:

> The concept expressed by w_1 as uttered by x_1 = the concept expressed by w_2 as uttered by x_2 iff the set of constitutive principles for w_1 for x_1 is identical to the set of constitutive principles for w_2 for x_2.

[5] For example, after introducing (Concept), Scharp writes: "Now we have concepts and we know exactly how we know things about them—by knowing things about *their* constitutive principles" (2019: 440, my emphasis).

It follows that, on this view, 'my concept of hoppiness' might be slightly different to 'your concept of hoppiness'. Scharp accepts this consequence and takes it to be unproblematic: speakers assume that they are expressing the same concept by a word until presented with evidence to the contrary (2019: 440); and different speakers can typically use their own concepts to interpret others with sufficient success to enable "quick and effective communication" (p. 441). For ease, I henceforth use "(Concept)" to denote this amended abstraction principle.

Now, (Concept) implies the existence of individual concepts. Consider:

(3) The concept expressed by "hoppy" as uttered by Mark = the concept expressed by "hoppy" as uttered by Mark iff the set of constitutive principles for "hoppy" for Mark is identical to the set of constitutive principles for "hoppy" for Mark.

This trivial instance of (Concept) implies the existence of 'my concept of hoppiness'. To see this, notice first that it is a biconditional. So, if it is true, it follows that its left-hand side has a truth value. The left-hand side has a truth value, though, only if "the concept expressed by 'hoppy' as uttered by Mark" has a denotation. So, (3) implies the existence of a concept, call it $\text{HOPPINESS}_{\text{MARK}}$, which is expressed by "hoppy" as uttered by me. On Scharp's view, we take the constitutive principles for $\text{HOPPINESS}_{\text{MARK}}$ to be precisely the constitutive principles for "hoppy" for me. I say that $\text{HOPPINESS}_{\text{MARK}}$ is *defined by abstraction* over "hoppy" and me.

We are now ready to state Scharp's answer to the grounding question.

> A concept, *c*, has constitutive principles, $P_1,\ldots,P_n$ in virtue of *c* being defined by abstraction over *w* and *x*, such that $P_1,\ldots,P_n$ are the constitutive principles for *w* for *x*.

Equivalently:

> A concept, *c*, has constitutive principles, $P_1,\ldots,P_n$ in virtue of *c* being defined by abstraction over *w* and *x*, where denials of $P_1,\ldots,P_n$ would lead *x* to initiate meaning reflection with respect to *w*.

Thus, for Scharp, the meaning reflection mode of conversation plays a central role in grounding claims about which principles are constitutive for a given concept.

From this perspective, Scharp's claim that 'TRUTH is inconsistent' must be interpreted as something of a generalization: there may be no unique 'concept of truth'. We can easily make Scharp's claim more precise. For each English-speaking person *x*, let TRUTH_x be the concept defined by abstraction over "true" and *x*. Then Scharp's claim might be restated along the following lines: for typical *x*, TRUTH_x is inconsistent. In particular: for typical *x*, each instance of (T-In) and (T-Out) is constitutive for "true" for *x*. For ease of discussion, however, I make the simplifying assumption that there is a single concept, TRUTH, such that the denial of any of truth's constitutive principles (in a conversation) triggers meaning reflection with respect to "true" (in that conversation).

Before proceeding, let me highlight an additional feature of Scharp's account of concepts that will be important below. According to Scharp, constitutive principles have *normative* import: they are "*justified* by virtue of their meanings" (2019: 442). In particular, speakers have a default and defeasible *entitlement* to constitutive principles; and there is a default and defeasible *impermissibility* to deny a constitutive principle. Scharp takes such normativity to play a role in *guiding* conversation:

> It is the warpings and tiltings of epistemic standards for various claims that is the contribution of constitutive principles. They act like the rails of a railroad track—guiding the direction of the conversation without dictating it. […] [C]onstitutive principles make it more difficult to take up certain positions and easier to take up other positions in a conversation. (2019: 443)

Thus, on this view, the inconsistency of TRUTH can direct conversation towards incoherence (perhaps like a converging pair of rails on a railroad). When discussing such sentences as λ, the constitutive principles of TRUTH guide us towards asserting an impermissible sentence—namely, a contradiction.

2.2 The conceptual engineering project

Scharp claims that the inconsistency of TRUTH makes it unsuitable for use in inquiry: as terminological precision is required in inquiry, we want to safeguard against being guided towards impermissible claims. The problem is particularly stark in the field of truth-conditional semantics:

when we try to provide a truth-conditional semantic theory for a fragment of natural language that contains a liar sentence, it is almost impossible to avoid contradiction. So Scharp engineers two replacement concepts, ASCENDING TRUTH and DESCENDING TRUTH, by, in effect, stipulating their constitutive principles. Scharp intends these replacements to play the explanatory roles typically afforded to TRUTH.

The guiding idea behind ASCENDING TRUTH and DESCENDING TRUTH is to split apart (T-In) and (T-Out). That is, each instance of the following unrestricted schemas is constitutive for ASCENDING TRUTH and DESCENDING TRUTH respectively:

(A-In) If ϕ, then $\langle\phi\rangle$ is ascending true.

(D-Out) If $\langle\phi\rangle$ is descending true, then ϕ.

For so-called *safe* sentences, such as those with no semantic vocabulary, ASCENDING TRUTH and DESCENDING TRUTH operate just like TRUTH: the converses of (A-In) and (D-Out) hold for such sentences, and thus the ascending and descending truth values of such sentences coincide. This contrasts with certain self-referential sentences, such as 'ascending liars' and 'descending liars':

λ_A λ_A is not ascending true.

λ_D λ_D is not descending true.

It is easy to show that both λ_A and λ_D are ascending *true* and descending *false*. For such sentences, the converses of (A-In) and (D-Out) do *not* hold.[6]

Now, the precise details of the proposal need not concern us here, but a few additional comments are in order. Firstly, the instances of (A-In) and (D-Out) do not *exhaust* the constitutive principles of ASCENDING TRUTH and DESCENDING TRUTH. Scharp introduces ASCENDING TRUTH and DESCENDING TRUTH by means of an axiomatic theory, ADT, which includes (A-In) and (D-Out) amongst other axiom schemas.[7] As I understand the proposal, each axiom of ADT is a constitutive

[6] E.g., while it is provable that $\langle\lambda_A$ is not ascending true$\rangle$ is ascending true, we cannot infer that λ_A is not ascending true.

[7] For a full list of axioms, see Scharp 2013: 154.

principle for ASCENDING TRUTH and/or DESCENDING TRUTH. Importantly, ADT is provably consistent, and is (I think) jointly consistent with facts such as that λ is ⟨λ is not true⟩, that λ_A is ⟨λ_A is not ascending true⟩, and that λ_D is ⟨λ_D is not descending true⟩.[8] That is, so defined, ASCENDING TRUTH and DESCENDING TRUTH appear to be *consistent* concepts.

Secondly, Scharp does not provide a 'background picture' of these concepts to underpin or illuminate his axiomatization.[9] The consequence is that, as things stand, it is difficult to get an intuitive grasp on what ascending truth and descending truth really are. Although this is potentially problematic (see Bacon 2019: 380–385), I only want to flag the issue here. At present, ASCENDING TRUTH and DESCENDING TRUTH can be understood to be technical concepts, whose most illuminating constitutive principles are the instances of (A-In) and (D-Out) respectively.

Finally, Scharp does not claim that *everyone* should start using ASCENDING TRUTH and DESCENDING TRUTH in place of TRUTH *all the time*. The claim is rather that: the inconsistency of TRUTH causes certain problems in inquiry; and, in these problematic cases, we should use ASCENDING TRUTH and DESCENDING TRUTH in place of TRUTH. Scharp provides an in-depth example: truth-conditional natural language semantics. In broad outline, the proposal is to replace 'truth conditions' with 'ascending truth conditions' *and* 'descending truth conditions'. Thus, for each sentence *s* of a fragment of natural language, rather than yielding a theorem of the form:

s is true iff *p*,

Scharp's semantic theory yields a pair of theorems of the form:

s is ascending true iff p_1,
s is descending true iff p_2.

For further details, see Scharp 2013: 207–215, 225–253. For critical discussion, see Pinder forthcoming-a.

[8] For Scharp's consistency proof, see his 2013: 178–187. Bacon (2019) also provides a simple model for ADT.

[9] Although see Scharp 2013: 200–203.

That, in broad outline, is Scharp's conceptual engineering project.

3. Tension

The two aspects of Scharp's view are in tension: the inconsistency theory includes an account of concepts that implies the failure of the conceptual engineering project.

Consider the conceptual engineering project. Scharp intends to engineer *consistent* concepts that are suitable for use in inquiry, and then to *use* those concepts to develop a new approach to natural language semantics. To achieve this, Scharp needs to do two things. On the one hand, he needs to stipulate the constitutive principles for ASCENDING TRUTH and DESCENDING TRUTH. It is only by stipulating the principles that Scharp can ensure that the concepts are consistent. On the other hand, Scharp needs to use "ascending true" and "descending true" to *express* ASCENDING TRUTH and DESCENDING TRUTH. This is the mechanism by which Scharp *uses* those concepts to build a semantic theory.[10] Drawing these points together, we can state a precondition of the success of Scharp's conceptual engineering project: that he is able to *stipulate* the constitutive principles of the concepts expressed by "ascending true" and "descending true".

Turn now to Scharp's account of concepts, which was developed as part of his inconsistency theory to answer the grounding question. On his account, a concept *c* has constitutive principle *P* only in virtue of *c*'s being defined by abstraction over a word, *w,* for which *P* is a constitutive principle.[11] And, importantly, whether or not *P* is a constitutive principle for *w* is an *empirical* matter, determined by whether or not the negation of *P* would trigger meaning reflection with respect to *w*. A direct consequence of these points is this: what the constitutive principles of the concepts expressed by "ascending true" and "descending true" are is an *empirical* matter.

[10] These two things—stipulating constitutive principles for his concepts, and expressing them—are not necessarily tasks to be completed sequentially. As an anonymous referee points out, stipulating constitutive principles for a concept may require one to use a word to express that concept. The exact mechanisms in play here deserve further study, but this is not a task I can undertake here.

[11] For readability, I temporarily drop the relativization to people.

We can now state the tension explicitly. Consider the following question:

(Q) What are the constitutive principles of the concepts expressed by "ascending true" and "descending true"?

It is a precondition of the success of Scharp's conceptual engineering project that he can *stipulate* an answer to (Q). But Scharp's account of concepts implies that (Q) is an *empirical* question. And we cannot stipulate answers to empirical questions. So Scharp's account of concepts implies that his conceptual engineering project will fail.

Is there an easy way to resolve the tension? There are two obvious strategies that one might consider, neither of which are promising.

First, Scharp might deny that his account of concepts implies that (Q) is straightforwardly an empirical question. He might claim: speakers are in control of when they initiate meaning reflection, so a speaker can *decide* what the constitutive principles for a given word are *for herself*. That is, Scharp might claim that he has the power to simply *decide* that he would initiate meaning reflection with respect to "ascending true" and "descending true" precisely whenever an interlocutor denies one of the axioms of ADT. If this is right, then he can *a fortiori* decide what the constitutive principles of the concepts he expresses by "ascending true" and "descending true" are.

This strategy does not solve the problem, however. Scharp can very well *intend* to initiate meaning reflection whenever an interlocutor denies one of the axioms of ADT. But whether or not he would *succeed* remains an *empirical* question. Suppose that, during an intense debate with Scharp about his view, I denied one of the axioms of ADT. It is quite possible that Scharp, whatever his prior intentions, would *not* respond by reflecting on what I meant by "ascending true" and "descending true"—rather, he might suspect that I had made a logical mistake, or that I had forgotten the relevant axiom of ADT. Neither Scharp nor anyone else can know, without empirical investigation, whether he would initiate meaning reflection in such a case. The point here is this: according to Scharp's inconsistency theory, constitutive principles are determined by which utterances *would in fact* trigger meaning reflection, regardless of one's intentions. And that remains an *empirical* question.

The second strategy for resolving the tension is to deny that the success of Scharp's conceptual engineering presupposes that he can stipulate an answer to (Q). Perhaps Scharp can accept that, in the end, it is an *empirical* question what the constitutive principles for the concepts expressed by "ascending true" and "descending true" are; and *a fortiori* that it is an empirical question what the constitutive principles for ASCENDING TRUTH and DESCENDING TRUTH are. The thought would be this: we cannot *stipulate* answers to empirical questions; but we can *influence* them. So, Scharp might seek to *influence* the occasions in which people initiate meaning reflection with respect to "ascending true" and "descending true"—and, *if* successful, *then* his conceptual engineering project would be successful.

An initial problem facing this strategy is that Scharp does not intend to be the *only* theorist using ASCENDING TRUTH and DESCENDING TRUTH. The goal is for "a natural language, like English" to be supplemented with "an ascending truth predicate and a descending truth predicate" so that, whenever relevant precision is required, "conversational participants can switch from talk of truth to talk of ascending truth and descending truth" (Scharp 2013: 275). To achieve *this* on the present strategy, Scharp would need to influence a vast number of speakers to initiate meaning reflection whenever their interlocutors denied an axiom of ADT. Such a task would, I think, be almost impossible to complete. That is, if Scharp accepts that (Q) is an empirical question, then his conceptual engineering project will almost certainly fail.[12]

A more fundamental problem, though, is that this strategy gives us the wrong success conditions for Scharp's conceptual engineering project. On this strategy, the success of the project depends on Scharp's ability to influence the occasions on which people would initiate meaning reflection. But consider the following scenario. Suppose that people start using "ascending true" and "descending true", and that they *try* to use those words in accordance with the axioms of ADT. Amongst these people, suppose, is a collection of truth-conditional semanticists who begin to publish

[12] Cappelen (2018) accepts that conceptual engineering (in general) is effectively impossible in practice, but that we should continue regardless. For responses, see e.g. Koch 2018 and Pinder 2018. As I doubt that Scharp shares Cappelen's stance, I put Cappelen's view aside herein.

research papers in which they provide 'ascending truth conditions' and 'descending truth conditions' in place of 'truth conditions'. In such a scenario, Scharp's conceptual engineering project has seemingly been hugely successful.

But suppose that, in addition, a general practice develops in which people who accidentally contradict axioms of ADT are corrected for *logical* mistakes, and that meaning reflection is *not* triggered. Does this additional supposition undermine the success of Scharp's project? On the present strategy for resolving the tension, the answer would be "yes": as speakers would not initiate meaning reflection when an axiom of ADT is denied, they would *not* count as using ASCENDING TRUTH and DESCENDING TRUTH, and so Scharp's project would count as having failed. But that is absurd: on the present scenario, speakers are *trying* to adhere to the axioms of ADT, and so they *should* count as using ASCENDING TRUTH and DESCENDING TRUTH in some relevant sense. Whether speakers initiate meaning reflection when an axiom is denied should be irrelevant. The success of Scharp's conceptual engineering project should not be thought to turn on obscure facts about the occasions on which speakers would initiate the meaning reflection mode of conversation.

4. Revising Scharp's account of concepts

Scharp's account of concepts, as it stands, implies the failure of his conceptual engineering project. I recommend that Scharp revises his account of concepts. One way that Scharp might proceed is by re-examining two key features of his account.

The first key feature of Scharp's account of concepts is its answer to the grounding question. Providing an answer is essential, because it justifies the claim that TRUTH is inconsistent. On Scharp's view, the constitutive principles for TRUTH are the sentences the denial of which would trigger meaning reflection for "true".

The second key feature is the claim that constitutive principles *guide* how we use words. This claim is important because it explains why the inconsistency of TRUTH is a problem: TRUTH, in virtue of its inconsistency, will sometimes guide us towards asserting falsehoods. If constitutive principles

did *not* connect up with our linguistic behaviour in some such way, then it would be far from obvious that the alleged inconsistency of TRUTH was a problem.

Let me re-examine these key features in turn, starting with Scharp's approach to the grounding question. Scharp answers that question by appealing to an abstraction principle, (Concept). Now, Scharp intends us to interpret this principle as defining *concepts*. But another interpretation is also available: we could, instead, interpret (Concept) as defining concept *expression*.

The distinction can be brought out by considering an unusual set of constitutive principles. Consider one of Scharp's examples of an inconsistent concept, RABLE, whose constitutive principles might be characterised thus:[13]

(4) all tables are rables.

(5) no red things are rables.

Let us suppose that, as a matter of fact, *nobody* initiates meaning reflection precisely when (4) or (5) is denied, with respect to "rable" or any other word. Now, consider the following question: Does there exist a concept, RABLE, whose constitutive principles are precisely (4) and (5)? If we take (Concept) to define *concepts*, the answer is *no*: concepts are nothing more than abstractions over words and speakers so, if (4) and (5) are not the constitutive principles for any word for any speaker, there is no concept constituted by the principles (4) and (5). In contrast, if we take (Concept) to merely define concept *expression*, then the answer is *yes*: concepts exist independently, and (Concept) defines only what it is for a concept to be expressed.

What, on this picture, are concepts? To answer this question, we might interpret 'constitutive principle' both literally and minimally. That is, we might identify concepts with collections of

[13] See Scharp 2013: 36. RABLE is inconsistent as (4) and (5) are jointly inconsistent with the fact that there are red tables.

principles: any collection of principles, whether expressed by a word or not, is a concept.[14,15] So construed, there *exists* a concept whose constitutive principles are precisely (4) and (5)—viz. the collection consisting of precisely (4) and (5)—and we can call that concept RABLE. And then, a speaker, *x*, *expresses* RABLE by "rable" just in case: whenever (and only whenever) faced with a denial of (4) or (5), *x* would initiate meaning reflection with respect to "rable".

Now, as noted, Scharp explicitly takes (Concept) to define concepts (2019: 440). However, this strikes me as a mistake. If (Concept) defines *concepts*, then the existence of ASCENDING TRUTH and DESCENDING TRUTH depends on there being speakers who initiate meaning reflection with respect to "ascending true" and "descending true" whenever axioms of ADT are denied—making the very existence of ASCENDING TRUTH and DESCENDING TRUTH an empirical question. This sits awkwardly with how Scharp thinks of his conceptual engineering project. He takes himself to have successfully engineered two consistent concepts, ASCENDING TRUTH and DESCENDING TRUTH, that he is now encouraging others to use. Making the existence of these concepts an empirical matter only confuses the issue.

Instead, I suggest Scharp take (Concept) to define concept *expression*. Then he can accept that ASCENDING TRUTH and DESCENDING TRUTH exist independently of words and speakers, along with many other expressed and unexpressed concepts. From this perspective, Scharp has not *created* ASCENDING TRUTH and DESCENDING TRUTH: collections of principles can neither be brought into, nor eliminated from, existence. Rather, Scharp has *picked them out* by specifying their constitutive principles. It is in this sense that he has 'engineered' ASCENDING TRUTH and DESCENDING TRUTH.

[14] Perhaps not *every* collection of principles should be deemed a concept. Perhaps the principles have to be related in a suitable way. For example, there may be no concept whose constitutive principles are *1+1=2* and *moons orbit planets*, even though the former might be constitutive of some concepts (e.g. NUMERICAL EQUALITY) and the latter of others (e.g. MOON). I put this issue aside, however.

[15] This is not a proposal for how to think of concepts *qua* constituents of thought, nor is the view a proposed competitor to (say) the theory theory of concepts or the prototype theory of concepts. Rather, this is a proposal for how to think about the principle-constituted entities that play a central role in Scharp's theory. I put aside the (important) question of whether such entities ultimately deserve to be called "concepts".

Had he specified different principles, ASCENDING TRUTH and DESCENDING TRUTH would still have existed, but Scharp would have picked out different concepts. What (Concept) tells us, from this perspective, is only whether speakers successfully *express* ASCENDING TRUTH and DESCENDING TRUTH when using "ascending true" and "descending true".

I have suggested that Scharp take (Concept) not to define *concepts*, but to define concept *expression*. I have given some initial motivation for this revision, but the principal reason for the suggestion is that it will allow Scharp to respond to the tension raised above. Before we can see how it does this, though, I must re-examine the second key feature of Scharp's account—that constitutive principles guide how we use words.

As explained above, Scharp highlights one way that constitutive principles guide how we use words: via meaning reflection. If a speaker *denies* a constitutive principle, then this triggers meaning reflection; and a speaker *accepting* a constitutive principle will appear *justified* in virtue of the meanings of the words involved. In such a case, in light of my comments above, we say that the speaker uses the word to *express* the concept.

Now, there are questions we might ask about the precise mechanisms in play here. In particular, it is unclear to me what *grounds* the normative nature of the constitutive principles, and how this normativity allows the principles to impact on linguistic behaviour. Regardless, for present purposes, let us assume that constitutive principles *can* guide how we use words in roughly the way Scharp envisages.

This assumption notwithstanding, there are *other* ways that constitutive principles can guide how we use words. Consider, for example, the sense in which (A-In) and (D-Out) might guide our use of "ascending true" and "descending true". In my case, at least, (A-In) and (D-Out) guide my use of those words in the sense that: I explicitly intend to adhere to Scharp's characterisation of ASCENDING TRUTH and DESCENDING TRUTH; I know that his characterisation of those concepts appeals principally to (A-In) and (D-Out); and so I intend to adhere to those principles. For example, I take it as given that

(6) if λ_A is not ascending true, then $\langle \lambda_A$ is not ascending true$\rangle$ is ascending true,

but I take its converse to be something to be proved or disproved. This is an explicit attitude that I take towards "ascending true" and "descending true", and it need have nothing to do with meaning reflection: if you deny (6), I am more likely to refer you to the relevant axiom of ADT than to question whether we mean the same thing by "ascending true".

In this way, constitutive principles can guide our use of words via an explicit intention. Importantly, the intention need not be directed at the constitutive principles themselves; neither (A-In) nor (D-Out) need be part of the *content* of my explicit intention for my use of "ascending true" and "descending true" to be guided in this way. I have an explicit intention *to use "ascending true" and "descending true" in accordance with Scharp's characterisation thereof*. As Scharp's characterisation principally involves (A-In) and (D-Out), those principles will guide my use of "ascending true" and "descending true" whenever I am acting on that intention with sufficient care. This may involve me recalling those principles from memory, referring back to the text, asking my interlocutor, or something else. Whatever the details, the constitutive principles count as guiding my use of the relevant words via an explicit intention.

Let me make two brief points about being guided by constitutive principles via an explicit intention. Firstly, there is nothing unusual about being guided in this way. When authors of an academic, legal or medical text explicitly define a word, they typically expect their readers to intentionally interpret that word according to the given definition. When teaching logic to students, we encourage them to try to follow the relevant rules of inference precisely—and we intend to do the same when constructing proofs in our own research.

Secondly, guidance via explicit intention cannot be reduced to guidance via meaning reflection. When an individual is guided via explicit intention, she *might* initiate meaning reflection when an interlocutor says something that contradicts the definition/rule of inference; but, equally, she might *not*. In many cases, she will assume that the interlocutor *intends* to use the word according to the definition/rule of inference, and that the interlocutor has simply made a mistake—a mistake which is easily remedied by pointing back to the relevant definition/rule of inference, but does *not* call for meaning reflection.

The suggestion, then, is this. Scharp should accept that there are multiple ways in which constitutive principles can guide the use of words. One way that constitutive principles might guide our use of words is via meaning reflection. Another way that constitutive principles might guide our use of words is via explicit intentions.[16]

Now, when a collection of constitutive principles guides our use of a word via meaning reflection, I said above that we use that word to *express* the concept constituted by those principles. In light of the present revision, however, we can distinguish between two *different* ways that a word might be said to express a concept. When a concept guides our use of a word via meaning reflection, let us now say that we use that word to *semantically-express* the concept—this is what is defined by (Concept). And when a concept guides our use of a word via explicit intention, let us say that we use the word to *speaker-express* the concept.[17]

Drawing the discussion in this section together, we obtain the following picture. Concepts are collections of principles. Some concepts guide how we use words. There are at least two ways that this can happen: via meaning reflection and via explicit intention. A concept guides a speaker's use of a word via meaning reflection just in case denials of constitutive principles (and nothing else) would lead the speaker to initiate meaning reflection with respect to that word. The abstraction principle (Concept) tells us that, in such cases, the concept is *semantically-expressed* by the word in question as uttered by that speaker. When a concept guides a speaker's use of a word via explicit intention, she intends to use that word in a certain way. She might intend to adhere to the constitutive principles of that concept directly, say if she intends *to use "ascending true" in accordance with (A-In)*; or she might have a more general intention, such as *to use "ascending true" in accordance with Scharp's characterisation thereof*. In such a case, the concept is *speaker-expressed* by the word in question, as uttered by the speaker.

Let us now apply this picture to the tension raised in §3.

[16] There may be other ways that constitutive principles can guide our use of words. I will not explore this possibility here.

[17] This terminology reflects the standard semantic/speaker meaning distinction. See e.g. Grice 1989. Cf. Pinder forthcoming-b.

5. Resolving the tension

How does TRUTH guide our use of "true"? It is unlikely that many speakers, when using "true", explicitly intend to adhere to instances of (T-In) and (T-Out), nor to those of other related principles (such as the T-schema). Most speakers are not even familiar with such principles. So, of the two options sketched in the previous section, it is more likely that TRUTH guides our use of "true" via meaning reflection than via explicit intention. Let us assume that TRUTH does indeed guide our use of "true" via meaning reflection.

How might ASCENDING TRUTH and DESCENDING TRUTH guide our use of "ascending true" and "descending true"? In contrast to "true", most users of "ascending true" and "descending true" are presumably guided by their explicit intentions to adopt Scharp's terminology, rather than by meaning reflection. So, of the two options sketched in the previous section, it is more likely that ASCENDING TRUTH and DESCENDING TRUTH guide our use of "ascending true" and "descending true" via explicit intention than via meaning reflection. Let us assume that those who use "ascending true" and "descending true" are indeed guided in this way.

This contrast between TRUTH on the one hand, and ASCENDING TRUTH and DESCENDING TRUTH on the other, allows us to resolve the tension raised in §3. When our use of a word is guided by a concept via meaning reflection, it is an *empirical* question what the constitutive principles of that concept are. For example, it is an *empirical* question whether Scharp is right that a denial of an instance of (T-In) will trigger meaning reflection for "true". In contrast, when our use of a word is guided by a concept via explicit intention, the relevant constitutive principles can be *stipulated*. For example, Scharp picked out ASCENDING TRUTH and DESCENDING TRUTH by *stipulating* that the axioms of ADT should guide use of "ascending true" and "descending true". If we are happy to co-ordinate with Scharp, then we do so by *intending* to adhere to those axioms when we use "ascending true" and "descending true". We then use "ascending true" and "descending true" to speaker-express ASCENDING TRUTH and DESCENDING TRUTH.

It is worth emphasising the need for care on these points. On the present picture, concepts *are* collections of principles, and so have their constitutive principles *essentially*. If TRUTH is the

particular concept that in fact guides our use of "true" then, on the present picture: TRUTH has its constitutive principles essentially; but we must discover those principles, and thus the identity of TRUTH, empirically. So it is an empirical matter whether TRUTH is inconsistent in the sense that: it is an empirical matter which concept guides our use of "true" and, thus, it is an empirical matter whether our use of "true" is guided (via meaning reflection) by an inconsistent concept.

Similar care should be taken with the claim that constitutive principles can be stipulated. If ASCENDING TRUTH and DESCENDING TRUTH are *particular* concepts, then they too have their constitutive principles essentially. But, nonetheless, Scharp *can* stipulate that he intends to use "ascending true" and "descending true" in accordance with particular constitutive principles; and, in doing so, Scharp is effectively stipulating *which* concepts are to guide his use of those words. One cannot stipulate what the constitutive principles of a *particular* concept are; but one *can* stipulate that one will use whichever concept has such-and-such constitutive principles.

So, on the present picture, there is a clear sense in which it is an *empirical* question whether TRUTH is inconsistent; and there is a clear sense in which Scharp is able to 'engineer' ASCENDING TRUTH and DESCENDING TRUTH, by stipulating *which* concepts will guide his use of "ascending true" and "descending true".

Recall that, as presented in §3, the tension turned on the following question:

(Q) What are the constitutive principles of the concepts expressed by "ascending true" and "descending true"?

As Scharp sets up his view: it is a precondition of the success of his conceptual engineering project that he can *stipulate* an answer to (Q); but Scharp's account of concepts implies that (Q) is an *empirical* question.

It is clear where, according to the picture I have sketched, this goes wrong: Scharp does not distinguish between *semantic*-expression and *speaker*-expression. When we do so, we see that there are (at least) two ways to interpret (Q):

(Q_1) What are the constitutive principles of the concepts *semantically*-expressed by "ascending true" and "descending true"?

(Q_2) What are the constitutive principles of the concepts *speaker*-expressed by "ascending true" and "descending true"?

Now, the claim that the inconsistency of TRUTH is an empirical matter implies that (Q_1) is an empirical question. On the other hand, the precondition of the success of Scharp's conceptual engineering project is that we can stipulate an answer to (Q_2).[18] But, of course, there is no tension between the claim that (Q_1) is an empirical question and the claim that we can stipulate an answer to (Q_2). If Scharp adopts the picture sketched herein, the tension facing his account dissipates.

6. Closing remarks

The tension I have been discussing arises due to two of Scharp's aims. First, he aims to diagnose a certain concept as deficient *a posteriori*. Second, he aims to stipulate improved replacements. At first sight, these two aims seem complementary: *find something broken, then mend it*. But care needs to be taken in spelling out how concepts can be subject both to *a posteriori* inquiry and stipulation. If care is *not* taken, then the aims end up in tension. This is just the problem that I raised in §3: Scharp spells out the empirical nature of concepts at some length, unintentionally ruling out the possibility of stipulating a replacement.

Other conceptual engineering projects have similar aims. One set of examples are the ameliorative projects defended by, for example, Haslanger (2012) and Manne (2017). These projects aim to highlight an *a posteriori* deficiency in everyday concepts such as WOMAN, WHITE, and MISOGYNY, which purportedly have problematic social consequences, and then to *stipulate* and defend improved, revisionary definitions.

[18] On this approach, then, conceptual engineering can be understood in terms of *speaker-meaning*. As I have defended the 'Speaker-Meaning Picture of Conceptual Engineering' at length elsewhere (Pinder forthcoming-b), I will not consider objections to the approach here.

Another set of examples are those that employ the method of explication, typically associated with Carnap (1962) and Quine (1960). On one way of characterising the method, it involves replacing an imprecise concept, the *explicandum*, with a more precise alternative, the *explicatum*. Now, the imprecision of the explicandum is typically taken to be an *a posteriori* matter. This is hinted at by both Carnap (1962: 3–5) and Quine (1960: 238–239), but the point has been emphasised more recently by Shepherd and Justus (2015) and Koch (2019), who argue that experimental methods should be used to help clarify the explicandum, and by Schupbach (2017), who argues that the explicatum's similarity to the explicandum should be tested empirically.[19] Nevertheless, the explicatum is typically introduced by means of a stipulated definition or stipulated rules of use. For one example, Schupbach (2017: 692) suggests that we define an explicatum, EXPLANATORY POWER OF HYPOTHESIS h OVER EVIDENCE e, thus:

$$\mathcal{E}(e,h) = \frac{Pr(h|e) - Pr(h|\neg e)}{Pr(h|e) + Pr(h|\neg e)}.$$

And Quine's (1960: 237) principal example of an explication is the mathematician's stipulated replacement of the ordinary notion of an ordered pair, effectively defined thus:

If $\langle x,y\rangle = \langle z,w\rangle$, then $x=z$ and $z=w$.

Explicators, then, are naturally construed as seeking to stipulate an improved replacement for a concept that has been diagnosed as deficient *a posteriori*.

In light of the foregoing, we should ask: how are ameliorators and explicators understanding concepts, such that concepts so understood are subject both to *a posteriori* inquiry and *stipulative* definition? The general lesson from the present paper is that care must be taken here. If care is *not* taken, then the two aspects of the project may end up in tension—the ameliorator and explicator may end up diagnosing a deficiency that, by their own lights, cannot be fixed.[20]

[19] I have objected to both of these projects elsewhere. See Pinder 2017a and Pinder 2017b respectively.

[20] Like Scharp, I doubt that ameliorators and explicators would endorse Cappelen's (2018) view that their projects are worthwhile despite being almost certain to fail. (See note 12.)

I see no reason why ameliorators and explicators could not, in principle, follow a strategy similar to that developed in §§4–5. On this strategy, ameliorators and explicators alike might firstly adopt a theory of concepts as *definitions* or *rules of use*, and then provide a pluralist theory of how concepts so-construed can *guide*, or are *expressed* by, our use of words. Whether ameliorators and explicators would be willing to adopt the *details* from §§4–5 is less clear—if nothing else, both the identification of concepts as collections of principles, and the role of meaning reflection in determining which word semantically-expresses which concept, will no doubt be controversial. Nonetheless, the broad structure of the strategy certainly appears to be available for ameliorators and explicators alike. I leave it to individual theorists to provide precise details of how we can make sense of the dual *a posteriori* and stipulative nature of their projects.[21]

References

Bacon, A. 2019. Scharp on replacing truth. *Inquiry* 62(4): 370–386.

Cappelen, H. 2018. *Fixing Language*. Oxford: Oxford University Press.

Carnap, R. 1962. *Logical Foundations of Probability (Second Edition)*. Chicago: The University of Chicago Press.

Eklund, M. 2002a. Deep inconsistency. *Australasian Journal of Philosophy* 80(3): 321–331.

——. 2002b. Inconsistent languages. *Philosophy and Phenomenological Research* 64(2): 251–275.

——. 2007. Meaning-constitutivity. *Inquiry* 50(6): 559–574.

Grice, H.P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Haslanger, S. 2012. *Resisting Reality*. Oxford: Oxford University Press.

Koch, S. 2018. The externalist challenge to conceptual engineering. *Synthese*. doi: 10.1007/s11229-018-02007-6

——. 2019. Carnapian explications, experimental philosophy, and fruitful concepts. *Inquiry* 62(6): 700–717.

[21] Thanks to an anonymous referee for helpful comments.

Manne, K. (2017). *Down Girl: The Logic of Misogyny*. Oxford: Oxford University Press.

Patterson, D. 2007. Inconsistency theories: the significance of semantic ascent. *Inquiry* 50(6): 575–589.

——. 2009. Inconsistency theories of semantic paradox. *Philosophy and Phenomenological Research* 79(2): 387–422.

Pinder, M. 2017a. Does experimental philosophy have a role to play in Carnapian explication? *Ratio* 30(4): 443–461.

——. 2017b. On Strawson's critique of explication as a method in philosophy. Synthese. *doi*: 10.1007/s11229-017-1614-6

——. 2018. The Austerity Framework and semantic normativity. *Inquiry*. doi: 10.1080/0020174X.2018.1557543

——. Forthcoming-a. *Not* wanted: On Scharp's solution to the liar. *Erkenntnis*. doi: 10.1007/s10670-019-00170-x

——. Forthcoming-b. Conceptual engineering, metasemantic externalism and speaker-meaning. *Mind*. doi: 10.1093/mind/fzz069

Quine, W.V.O. 1960. *Word and Object*. The MIT Press.

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

——. 2019. Replies to Bacon, Eklund, and Greenough on Replacing Truth. *Inquiry* 62(4): 422–475.

Schupbach, J. 2017. Experimental explication. *Philosophy and Phenomenological Research* 94(3): 672–710.

Shepherd, J., & Justus, J. 2015. X-phi and Carnapian explication. *Erkenntnis* 80(2): 381–402.