The Descriptive And Normative Semantics Of Social Media

Abstract

I consider the descriptive and normative semantics of social media, focusing on twitter. The descriptive part of the paper uses tools from formal semantics to ask what the meaning of a tweet is, suggesting that tweets are importantly *reception sensitive*: part of their content is information about how they were received, in the form of representations of how many retweets, replies, and likes they got. The normative part of the paper asks whether this is a good content for a tweet to have, and uses discussion of this question to present an answer to a worry about normative semantic projects to the effect that they are doomed to failure because even if meanings are bad, there's nothing we can do about it because we can't change meaning. I conclude that, at least in the case of social media, meaning change is possible, because meaning is determined by the platform, and the platform is controlled by a small handful of engineers.

Keywords: social media, formal semantics, conceptual engineering, meaning, internet

1. Introduction

The platforms we inhabit on social media, via the interfaces they use, partly determine the semantic content (what is said; the proposition expressed) of what we say on social media. In tweeting, for example, part of what you say is determined not by you, but by the interface. This is bad.

This paper is devoted to spelling out the above argument. It has various moving parts, and so it'll be helpful to begin by foreshadowing the structure of the argument, and thus of the paper, in some detail.

I will argue firstly that the notion of semantic content, and with it the range of tools and styles of arguments we use when discussing semantic content, can be applied to social media communication. Then I'll make the case that the semantic content of a tweet is other than we might think. In particular, a tweet is what I'll call *reception sensitive*: part of the semantic content of a tweet is information about how that tweet was received by other users of the website.

I'll present an argument for this, but I can make the point quickly and intuitively. Open up twitter, and look at a tweet. In addition to the username, text, and time and date of the tweet, you will see icons and numbers that tell you how many replies that tweet received, and how many times it was retweeted and liked. Call those last three things the *reception information*. My claim is that if we apply the notion of semantic content as applied in formal semantics to tweets, we should say that reception information is part of the tweet.

That's the first part of the essay: using tools from formal semantics to get clear about the semantics of twitter. This is a descriptive project, not much different, in my opinion, from giving theories of, say, epistemic modals.

But recently a more normative strain of thinking about language has arisen. The field of conceptual engineering is motivated by concerns not about how language is, but how it should be, and by the desire, if language is suboptimal, that it be improved. And that, latter, question brings in its train further questions: how can we improve language, especially in light of the fact that many of us think that language is a social artifact not under the control of any of us (let alone academic philosophers!)? Herman Cappelen (2018) has argued that this fact--that the metasemantics of the languages we speak are out of our control--means that there is no clear path to carrying out a conceptual engineering project. We can think and talk and care about it, and even maybe do some small things to improve the situation, but that's likely to be no more effective than other small scale normative activity: it's like signing a petition against climate change.

The second part of the paper considers these questions. It suggests there should be very real concerns that the sorts of content we express on social media are suboptimal. We can argue for this both a priori and a posteriori: a priori, if we grant that part of what we communicate is determined by the interface, and that the interface is designed in part to keep us using the site as much as possible in the service of giving away our data and keeping eyeballs on ads, we should worry that the designers of the interface don't have optimal communication as one of their goals. A posteriori, it might just seem straightforwardly bad that you can't see a tweet without also seeing its reception. Seeing how others reacted to it will probably change how you react to it.

The consequence of this is that we should be wont to consider the second question: can we change it? Then Cappelen's worry becomes pertinent. If the metasemantic facts are out of our control, then it seems we can't. We're stuck with bad languages on social media. Since we spend all our time on social media, that's not a heartening conclusion.

But, and this is the second big point of the paper, arguably the metasemantics of social media communication are suitably different from the metasemantics of face to face communication such that even if the latter are, for externalism-based considerations, immutable, it doesn't mean the former are. The former readily are: all we need is to get the ear of twitter engineers; or, indeed, build our own alternative platforms. The crucial point to realize is that the at least one aspect of the metasemantics of social media communication is determined by the interface, and so by a handful of engineers. It's an engineer's decision that made reception information part of tweets, and it's an engineer's decision to remove it, and engineers, arguably, are more under our control than causal chains, or joints of nature, of the chaotic use of millions of people that the externalists tell us determine the meanings of what we say.

There's quite a lot to get through, and because I would like this paper to be read hopefully by people outside of philosophy, I am going to work from the bottom up, explaining the key notions. In the first section, I introduce the notion of semantic content that will be used throughout. I also put forward and briefly defend some central assumptions I will be making. In the second section, I consider social media language and make the argument above that on social media semantic content contains reception information. In the third, we're back to exposition, as I set out the conceptual engineering project, and in the fourth I consider the application of conceptual engineering to social media content.

2. The Theory Of Meaning

Words mean things. But what does *that* mean? What is it for an expression to have meaning?

Faced head on, such a question seems daunting. We might attempt introspection, thinking we know, in some intuitive sense, what meanings are. But do we? Appealing to our folk conception of meaning might lead one to think that meanings are dictionary entries, or again images or thoughts that words call to mind. But neither of these thoughts lead anywhere: to say that the meaning of a word is what the dictionary tells us is to say that it is more words, an arguably unhelpful regress, while saying that meanings are mental images, as has long been recognized (e.g. Frege 1884), can't capture the fact that meaning is public and shareable.

We might then try to try to zoom out, and bring to bear academic tools from evolutionary theory, psychology, game theory, or any of a number of other disciplines. We might try to explain how language arose and how it relates to more primitive signalling systems, or what goes on in the mind or the brain when using language. On this story, meaning would be understood in terms of, and reducible to, some separate theory.

A famous moment in 20th century intellectual history convinced many that this approach, at least as concerns syntax, was no good. Noam Chomsky (1959), reviewing B.F. Skinner's *Verbal Behaviour*, demolished the idea that the behaviourist psychology in vogue at the time was apt for explaining language, and in work around the same period showed how examining syntax closely, using formal tools from logic and computation theory, could yield explanatory theories of how language works.

Formal semantics, as I understand it, takes the same tack. It is based on the idea that there are rules about meaning, rules which neither we intuitively understand but nor can be cashed out in terms of other disciplines' theories, but which we can discover to yield elegant formal theories of meaning. My aim here is to show the first stages in the development of such theories, which will give us both a working grasp of the notion of meaning, and the style of argumentation employed by formal semantics. This, in turn, will give us tools to examine social media communication.

The central rule of formal semantics concerns the dependence on the meaning of its parts of the meaning of a complex expressions. To see this, consider a mundane sentence:

(1) Barack seldom golfed

Here are three obvious things about this: it has a meaning. It is composed of various words. And those words themselves also have a meaning.

But not only that: the meanings of the words that compose it, together with the way they are composed, somehow determines the meaning of the whole thing.

The whole thing says something about the sporting activities of the former president, and it does so by using words that talk about the former president and about sporting activities. This is an obvious but fundamental and deep fact, and it furnishes us with our first rule of the theory of meaning

> **Compositionality.** The meaning of a complex expression is determined by the meanings of its parts, and how they are composed.

This principle should seem intuitive to you: think of some sentences, and you should become convinced it's at least on the right track (spelling out the principle in detail is difficult and has generated a large amount of forbiddingly technical literature, but, thankfully, we can ignore that question for our purposes). Natural languages obey--and, you might think, although I won't make that case here--*couldn't not* obey the principle. At least some are clearly ruled out. It's hard to conceive of a language in which 'Obama', 'seldom', and 'golfed' mean what they do in English but the sentence meant 'It always snows in Iceland'.

Now armed with a fixed point, we can begin to ask more questions. For example: can we use Compositionality to help explain what meanings actually are? What, for example, does 'Obama' mean?

Well, consider this

(2) Barack seldom golfed

Note that, at the very least, what this means is much closer to what (1) means than 'It always snows in iceland' is. It's also much closer to what (1) means than, say 'Carter seldom golfed' is.

Given this, we might be tempted to say the ‘much closer to’ of the previous paragraph is in fact ‘the same as’, and that 'Barack' and 'Obama' mean the same.

You don't have to buy this judgement. You might think there are important shades of meaning with respect to which ‘Obama’ and ‘Barack’ disagree. But even if you do think that, it doesn't seem unreasonable to say that there's a concept of meaning worth exploring according to which the two mean the same. So let's explore it, and see if it's fruitful. Spoiler alert: it is.

We can note one important property that ‘Obama’ and ‘Barack’ both uncontroversially share: they both name or refer to one and the same object in the world, a particular person. Based on this, we might tentatively suggest the following as a second principle:

> **Name Meaning**. The meaning of a name is its bearer.

The problem with this, as stated, is that it's not general enough--science aims at generalizations, and not just at saying how things stand with particular things. It would be nice to have one principle that tells us what each of sentences, verbs, adjectives, etc., and not just names, mean. Can we find a more adequate generalization that captures the same idea as the particular principle we just gave for names?

Well, note this: one prima facie plausible and interesting thing about 'Barack' and 'Obama' is that if you replace one for the other in many types of sentence, you'll always wind up with something with the same truth value as before. Say 'Obama seldom golfed' is true. Then you can be guaranteed that 'Barack seldom golfed' is also true. And say ‘Obama always flew’ is false. Then you can be guaranteed that ‘Barack always flew’ is also false.

Based on this observation, we might propose the following more general principle:

> **Substitution.** Switching one expression for another always results in sentences with the same truth value iff the two expressions mean the same.

Note that this seems to work well for 'golfed'. If we switch it for the verb phrase 'played golf', we are guaranteed preservation of truth value, and it does seem, intuitively, that the two mean the same.

It seems like we might be getting somewhere. But now consider:

(3) Obama seldom golfed and Carter seldom swam

Assume this is true. Then note the following: if we replace 'Obama seldom golfed' with any other true sentence, we get a true sentence:

(4) Snow is white and Carter seldom swam

This might seem disastrous: all sentences with the same truth value have the same meaning!? But let's explore it some more--it might be that, although unintuitive, it helps.

And help it does. If we're looking for an object for sentences to stand for, and the important thing seems to be their truth value, then we could say that the meaning of a sentence is simply a truth value, where that is some weird abstract object. There are two truth values, True and False, and every sentence means one or the other.

Okay, so that's sounding weirder. But a handful of theoretically interesting consequences follow from this fact, and these facts should cause us to be more receptive to the possibility we're on the right lines.

For example, one question we left outstanding was what sort of thing verbs meant. But now we can note that Compositionality tells us the meaning of the sentence, a truth value, must be determined by the meaning of its parts, which is to say an object and whatever the meaning of verbs are. Working back, a verb meaning is something that combines with a object to yield a truth value.

Something that combines with one thing to yield another has a mathematical interpretation: a function. A function takes something in and returns something. We can then say that the meaning of a verb is a function, and in particular that function that maps an object to True iff that object does the thing in question. ‘Golfed’ is a function that maps an object to True provided that object golfed.

Taking truth values as sentence meanings, if a bit weird, leads to insight about verb meaning, which recommends it.

Here's another interesting consequence. Consider again:

(5) Obama golfed and Carter swam

Our compositionality principle--as well as intuition--nudges us towards assigning a meaning for 'and'. But given the idea that meanings can be functions, one quickly suggests itself. ‘And’ could be a function that takes a pair of truth values to another truth value, and in particular takes them to true iff both are true.

We could go on: we could give meanings for the adverb ‘seldom’, just about, that would obey the principles set above. And the plausibility of this whole interconnected skein of ideas: of compositionality, truth preservation, and functional semantics, and in particular how, although they seem like separate ideas, seem to illuminate each other, provide, I think, reasonable evidence that the theory we’ve been developing is latching on to something in linguistic reality.

2.1 Embedding Arguments

There's at least one point in the above where you are probably unsure--the meaning of sentences. It seems pretty weird, pleasing theoretical consequences aside, to say that all true sentences mean the same things, as do all false sentences.

We can make it seem weirder by introducing a central part of the methodology of formal semantics, one which will drive much of the later work on twitter, namely *embedding arguments*. In embedding arguments, we test our analysis of a bit of language by making sure that that analysis remains plausible when we consider more complicated environments in which that bit of language can occur.

Here's a simple sort of example to see what I mean. An erstwhile popular analysis of moral, aesthetic, and metaphysical discourse, associated with the logical positivists (e.g. Ayer 1936) had it that such language was meaningless. In saying

(6)Murder is wrong
(7) The Simpsons is amazing
(8) The universe is gunky

We are not, in fact, actually putting forward a claim. Rather, it's as if we were cheering or booing murder or the Simpsons--we are just expressing our attitudes, how we feel, about the things in question (how this would apply to a sentence like (8) is less clear, which is why I'm sneakily just going to dodge the question). An apt analysis of such sentences, at least as often presented, is as so: 'Boo, murder!', or 'Yay, The Simpsons'.

An embedding argument shows a problem with this analysis. Consider

(9) If murder is wrong, then you shouldn't kill people

This is meaningful, and indeed true. But if 'Murder is wrong' is meaningless, then it should be as meaningful as the following, which is rendered meaningless by having something meaningless in the antecedent of the conditional:

(10) If mostly people dancing, then you shouldn't kill people

The point is, by embedding the analysandum ('murder is wrong') in a complex environment (the if-then conditional) and seeing that the analysis yields the wrong results, we can conclude the analysis is no good (maybe: recent work, beginning with Schroeder 2008 have proposed sophisticated theories to deal with this problem).

Embedding arguments are ubiquitous in philosophy of language, and we will use one later. But relevantly, we can use one now to show that our truth value analysis of sentence meaning is no good. Consider:

(11) Necessarily, 2+2=4
(12) Necessarily, Obama seldom golfed

These two sentences have different truth values, and since one is got by replacing '2+2=4' and 'Obama seldom golfed', by Substitution it follows that those two sentences don't have the same meaning, but since they do have the same truth value, it follows that sentence meanings aren't truth values. An embedding argument shows our first-pass theory of meaning was wrong.

But that doesn't mean all the above was a waste of time. Because we now have the basic structure of the theory in hand, we can ask how the modify it to yield the right results. We need to assign a meaning to sentences consistent with the data presented above, and one that--hopefully--preserves as much of our nice seeming theory. That can be done, but I won't do it here (the interested reader could consult the textbooks Heim and Kratzer (1998) and Heim and Von Fintel (ms)).

Let's sum up. We started with a big and baffling question: what is meaning? And we edged our way to an answer that treats the theory of meaning as sui generis, subject to its own laws. And then we showed how to work out those laws, how they illuminate the nature of meaning, and one sort of argument we can use to test our theories. In the following section, with this theory of meaning in mind, I want to turn to social media communication, and see if we can shed light on it.

3. Twitter As A Formal Language

Here's a question as interesting-sounding and as vague as the one with which we began: what effect does social media have on how we communicate?

It's hard to answer because it's so big, and my strategy will be to try to make it smaller and more precise by asking what the tools of formal semantics, introduced in the previous section, can tell us about social media communication.

I will focus on twitter; some conclusions perhaps generalize, others don't, and for other platforms there will be similar truths in the same vein, which I leave for other work or others' work.

Twitter is a microblogging platform. Users tweet text messages of up to 280 characters, and which can include pictures, videos, and gifs. Users follow other users, an asymmetrical relation, and users can interact with a tweet by responding to it, or by retweeting it, which causes the tweet to be shown by those who follow the retweeter, or by liking it, which doesn't display the tweet on the user's timeline.

There is much to complain about. The short format tends against nuance and long replies. Differing follower counts lead to some having a bigger platform than others without always deserving it. That accounts are easily and anonymously made encourages armies of trolls and bots. Trump, plausibly, was helped by its existence. And the whole thing relies on people giving their time and data producing and interacting with content for the enrichment of the platform owners.

All of these things provoke interesting questions in pragmatics, various parts of epistemology, ethics, and more, and indeed the burgeoning literature on fake news shows that academia is interested in them.

My goal, as already stated, is to look at the formal semantics of twitter speech. Initially, this might seem like a very barren area of theorizing. After all, isn't the language of twitter simply English (or whatever natural language the tweeter happens to speak), and so accordingly won't the formal semantics of twitter just be the formal semantics of the tweeter's spoken language?

The central claim of this section is that the answer to this question is no. What we can call Twitter English is different from normal English. The reason it is so is because, part of the meaning--the basic, truth-conditional, compositional meaning introduced above--is information about how the tweet was received. This information consists in the number of replies a tweet received, as well as how many times and by whom it was favourited. The proposition expressed by a tweet, I claim, contains that information.

English doesn't contain that information. If I am see Theresa May give a speech, I learn nothing about its reception--about whether people like it or disliked it, whether they engaged with it or ignored it. I hear something like

(13) Brexit means brexit. Brexit will make us all better off.

I do not hear something like

(14) Brexit means brexit$_{\text{most people were baffled}}$. Brexit will make us all better off$_{\text{people snorted with derision}}$

Where the subscripted text, somehow, was automatically part of the sentence and encoded reception information.

By contrast, if I read Theresa May tweeting those sentences, I *do* see how it was received. It's part of the tweet as displayed--down at the bottom. A tweet without that information is not a tweet. The rules of twitter--the software that runs the platform--just don't allow for that. Saying something without reception information being included is impossible.

3.1 Arguments That Reception Information Is Part Of Semantic Content

This is the first argument for the claim that reception information is part of a tweet. It could be viewed, roughly, as something like a syntactic argument. It's a condition on a tweet's being well-formed that it contain reception information. Anything of necessity part of a piece of language contributes content to that piece of language.

Actually, this needs a bit of precisification. The reason for this is that the second premise doesn't always hold. In non-pro-drop languages like English, syntax sometimes requires things that receive no semantic interpretation. Thus to say it's raining, I must, of necessity, give the verb 'rain' a subject term 'it', even though it's widely agreed that this so-called expletive 'it' doesn't contribute anything to the meaning of sentences in which it occurs.

The following principle seems better: anything of necessity part of a piece of language, which has a semantic interpretation, contributes that interpretation to the piece of language. This seems, I think, like a more solid principle, and it's that on which I base argument one for the claim that reception information is part of the content of a tweet.

Argument two is an embedding argument. We've already seen them above, but recall an important feature of them: one way to test whether something is part of the semantic content of an expression, as opposed, perhaps, to a feature of the pragmatics, is to see if that content interacts with embedders.

Open twitter and look at a retweet. At the top, in small letters, it says who retweeted it into your timeline. My claim is that this text functions as an embedding environment, just as conditionals and necessity operators do. And what follows it is what is embedded. But note that what is embedded is the original tweet *along with the original tweet's replies numbers, likes, and retweets*. What that means is that the content of the retweet*ing* tweet makes use of that information, the reception information in the retweet*ed* tweet. Now here is a principle:

> If a piece of information associated with an embedded expression is necessary to determining the content of the embedding expression, that information is semantic content.

This is an independently plausible principle, which can be seen at work in some of the most famous embedding arguments. Thus consider Kripke's (1980) modal argument against descriptivism.

(15) It could have been that Nixon won the election

(16) It could have been that the winner of the election lost the election

Now, here's a possible analysis of 'the president'. It is, like the names we started with, an expression that simply stands for an object. The descriptive meaning is not part of the content.

The embedding argument gets its force from excluding this possibility (Kripke's actual argument was targeting an analysis of names, not descriptions, but it gets its force from the implicit theory of descriptions discussed here). We can't make sense of the idea that the descriptive meaning is not part of the content because that meaning interacts with the embedding modal operator, which evaluates the descriptive content relative to different possible worlds.

So, I claim, in this case. What one retweets includes reception information. Were that information not part of the semantic content, it couldn't interact with the retweeting environment (instead, that reception information would get lost as the composition procedure went forward. That yields the testable but falsified prediction that, since all tweets must have reception information, the reception information associated with a retweet would be the reception information of that retweet. That's a very real possibility; it's just not what is in fact the case.)

Treating retweeting as an embedding operator gives us, then, strong reason, in light of the fact that such an operator interacts with reception information, to say that that reception information is part of the content of a tweet.

Here's a third argument. If, as I have suggested, reception information is part of the compositional semantic content of the tweet, and if compositional semantic content at least loosely tracks what is said, then we should expect this: say we want to depict someone's saying something. Then, if what they say involves reception information, we should expect the depiction information to contain it also.

This sounds a bit weird and will be helped with an example. It would be very weird, if we wanted to fictionally depict someone's saying something in a play, if we were to leave out important parts of what they said. If we wanted to depict Trump saying he will build a wall and make Mexico pay for it, it would be weird to leave out obligatory parts of what he said--say, the verb ending, as so

(17) Trump say he build wall and make Mexico pay for it

And this is what we see. Increasingly, when people want to tweet in another voice, they write in the reception information (admittedly, they do so in the service of pointing out stuff about that reception information). Thus consider the text of this[1] tweet:

> Uhhhh date someone that you can *looks at scribbles on hand* do things with?
>
> 3k Replies | 7k Retweets | 15k Favs

Or again of this[2] one

> Girl: I just want to nap for 2 days, have 8 corgis, and not live in Utah.
>
> Retweets: 10.1 K Favs: 16.4 K

In both cases (which can be multiplied, but it's very hard to search for) the tweeters include, in the text of their tweet, reception information about the (fictitious) tweet. The explanation for this, I claim, is that they realize that the reception information is part of the content of the tweet.

3.2 Objections

The view I've put forward deserves serious consideration. Like most interesting views, though, it is subject to some objections. I choose to list them here both in the spirit of honesty but also because doing so establishes the metagoal of the paper, which is to make people take seriously the formal semantics of online communication.

Let's start with a not so serious objection. I claim that '[name] retweeted' forms an operator as part of a tweet. But then what is the truth value of such a tweet? One might be tempted, considering the analogy with conditionals and modals as embedding environments, to think that retweeting tweets have truth values, but then it seems one is compelled to say retweets are always true, which seems like a weird and unintuitive verdict. More comfortable is to say that they are non-truth-apt, but then it becomes less clear that we can analogize tweets to sentences, which are truth apt.

[1] https://twitter.com/Dakotalogy/status/998033416724639744

[2] https://twitter.com/chocomilk_7/status/883794859986898944

I deny that we should assign truth values to retweets. There are several ways to make such a denial. I could say that they are instances of non-sentential assertion. Twitter is awash with such assertions. More radically, I could say that it's not the business of a tweet as a whole to put forward something as true or false. It's sometimes the business of the embedded natural language text (but not always) but not the point of the whole tweet. A tweet is something sui generis: a form of language the main moves of which don't involve putting forward a claim (though they can typically have parts which do so). This is a defensible position, although it would require more work than I intend to do here. So consider this a promissory note.

For a second objection, can note the phenomenon of quote tweets. In a quote tweet, someone's tweet appears in a box in the tweeter's tweet, along with text from the tweeter. Quote tweets are used primarily to pass comment on another's tweet, not to just straightforwardly republish it as retweets do. The crucial thing is that in the box in which the other person's tweet appears, the reception information doesn't. If we want to treat quote tweets as embeddings, then this would undercut my claim that, because reception information is preserved in embeddings it is part of the semantic content of a retweeted tweet.

I think the thing to say is just that what is quoted isn't the full tweet: quote tweets are a form of embedding in which information is lost. That's hardly a complete novelty. Indeed, quotation itself can be seen as a form of information-losing embedding. When I say

(18) Jan said 'You come here now'

By reporting Jan's words, rather than what she meant by them, we lose information about what she did mean by them. Quotation allows us to report the words and not the meaning; similarly, quote tweets allow us to report the words and not the reception information.

A third objection is this. Say I retweet a tweet and then you do. Pretheoretically, we retweet the same tweet. But not on my analysis: because you retweet it after me, the tweet you retweet will have at least one more retweet than the tweet I do. Different content, different tweets, and so we don't retweet the same tweet.

This is a more troublesome objection, and I think properly to do justice to it requires more space than would be profitable to spend. But there are preliminary things to do. It's a familiar point from the language of reporting others' speech that our reporting practices are flexible. Cappelen and Lepore (1997) point to cases in which one can truly report a person as saying something by using a sentence notably different from the one they in fact used. If Carrie says 'On my birthday, I bought a pair of cute $700 heels from Sach's' I can report her utterance as so:

(19) Carrie said she bought an expensive pair of shoes on Tuesday

What Carrie literally said and what she is reported as saying are markedly different. But the report is fine. The motto is that one can count as reporting what a person says even if your report diverges from the actual content of what they say. It's sufficient that your report be close enough. Then my claim is that retweeting works the same way: strict identity of tweet is not required us to judge that two people retweeted the same tweet.

Let me turn to the final. I have claimed that reception information is part of the content of a tweet, and a crucial part of that was that retweets are tweets. But following that logic would surely demand that the reception information of the retweeting tweet be part of the retweeting tweet. But it isn't. In a sense it seems that what's happening is that the reception information of the retweeted tweet overrides that of the retweeted tweet. There is one slot in a tweet for reception information and that information, in retweets, is obligatorily the reception information of the retweeted tweet.

The objection, then, is: I claimed that the reception information of a tweet is part of its content, but that's not true, because retweets do not contain their reception information, and so retweets aren't tweets, and so the embedding argument doesn't get off the ground.

I think the response to this is just to modify my view slightly, and indeed that doing so is informative. We say that tweets must contain some reception information, although not necessarily the reception information of the tweet itself (in the case of retweets). This leads us to a position on which what represents the reception information is something like an obligatory syntactic part of a tweet which can receive varying semantic interpretations. This seems like a reasonable position; I think there must be some interesting natural language analogue of this, but I can't quite think of it right now. (The closest analogue I can think of is sequence of tense phenomena, whereby a verb's syntactically obligatory tense gets ignored because of a higher tense expression. But the parallel isn't quite perfect.)

Let me end by reiterating the meta point: I have been considering objections just as I would consider objections to any formal semantic theory, by looking closely at the data, and seeing whether our theoretical tools are apt for them. If you have found the discussion in this section useful, then I hope it's made you more on board with the idea that twitter is an apt object of serious semantic theorizing.

3.3 Consequences

Overall, then, I think a reasonable case can be made that the meaning of a tweet includes reception information. The case is not watertight, but cases in philosophy seldom are. In this section, I want to consider semantic, metasemantic, and socio-political consequences of what has come before, which will set up the second part of the paper.

The conclusion is important, for several reasons. For one, if I am right, twitter communication is interestingly new. The path of formal semantics has been one of increasing complexification, the first stages of which we saw. We complexify when we encounter intensional phenomena (as we did), or context sensitive phenomena (Kaplan 1989), or presupposition (for which see, e.g., Beaver 2011), or assessment sensitivity (e.g. MacFarlane 2014). Twitter involves, I think, a further complexification, and thus induces us to rethink semantics.

I won't pursue this question here, because, although I think it's interesting, I have enough to be getting on with (I will just say that my hunch is that something like Andy Egan's assessment sensitive utterance bombs (2009) are probably the theoretical framework that would yield the smoothest account of the fact that tweets contain reception information. An important part of a completed theory would need to take into account that the reception sensitivity of tweets is cumulative. Each person who retweets the tweet, in a sense, changes the content that gets taken up by any subsequent retweeter. This yields a new and interesting view roughly in the space of assessment sensitive semantic theories.).

There are also metasemantic questions: questions as to what determines what our expressions mean. The metasemantics of social media, if I am right, are notably different from normal externalist metasemantics. According to normal externalist metasemantics, meanings just ain't in the head, but are instead determined by the world. Content is external. On social media, content is external in an even more extreme sense. On the standard externalist semantics, you can at least chose the words, even if you can't choose what they mean. On twitter, however, there are certain mandatory parts of content over which you have no control. Tweet, and you thereby express a content that contains information about its own reception. It doesn't matter if you don't want to--if you'd rather, say, just have people attend to your words (or your cat pics or Michael Scott gifs), well, too bad. It's built in to the nature of tweeting that you express that content. What you mean is determined in part by the twitter platform. This is a new sort of metasemantic externalism; and this fact will later be used to shed light on conceptual engineering.

Finally, this latter fact should cause some worries. Let's grant that the metasemantics of social media communication is determined in part by the platform. And then note that twitter is a business, and as such its aim is to be profitable. Its aim is not to help us express the best contents that we can express. Plausibly, profit maximization and optimal communication pull in different directions. So there's an a priori worry that it's a bad thing to let what we say get routed through platforms in this way. And if we grant that, then we might naturally ask what the best sort of content we might express using social media might look like.

But once we're in the business of assessing communicational devices as good or bad, we have moved from descriptive semantics to normative semantics: from linguistic modelling to conceptual engineering. In the next, expository, section, I briefly outline conceptual engineering before, in the following one, returning to assessing social media communication and seeing how we might engineer it to make it better.

4. Conceptual Engineering

In this section, I motivate, and outline a central problem for the project of, conceptual engineering. Various writers, with various motivations and various terminology, have talked about this topic, and rather than attempting a synopsis, I will just focus on the presentation found in the work of Herman Cappelen (2018, forthcoming), which anyway the material in this paper is responding to.

Some languages are better than others. The variant of English we speak today is better than the variant of English we spoke in 1950 because today's English contains the terms 'sexual harassment' and 'post-partum depression', because the concepts these terms stand for are important for many people's (harassed people and harassers and bosses, recently pregnant women and those close to them) understanding of the world and because having a term that stands for a concept is one of the best ways of making a concept salient to a person. Languages are delivery mechanisms for contents, and contents determine how we conceive of and understand the world, and that in turn determines how we act (this is a point emphasised in Plunkett and Sundell 2013).

If we realize that English today is better--in the respects just mentioned, it may be worse than others--then we should also realize that it's wildly implausible to think that English is as good as it can be, and so, given how important words are as bearers of meanings, we should think it's an important project to try to improve English. This, more or less, is the master argument found in Cappelen (forthcoming).

The master argument, when presented like that, is compelling. But there are also very strong arguments against the very possibility of conceptual engineering. Some relate to the thought that one can change the meaning of particular expressions. A reasonable view would have it that a word is in part individuated by its meaning such that if you change the meaning, you change the word, and so it's confusing not to change the lexical realization of the word.

The one that I'm concerned with has it that meanings aren't in our control, and so even if we realize that our language could be better, there is nothing we can do to make it better.

In a bit more detail, this argument starts from the premise that externalism about meaning is true. According to externalism, what determines the meanings of words is not, for the most part, something under our control.

Externalism comes in a variety of forms. Kripke argued that the meaning of names is determined by certain causal relations between the users of a name and the original bearer.
On this story, there is a baptism whereby a name is initially bestowed upon an object, typically by someone pointing to it and saying 'I call you [name]'. Then that person uses the name to others, and and even if those others haven't directly encountered the named object, because they've learnt the name from someone who has, the name in their mouth comes to stand for that named object. This story continues: the others spread the name yet further, and new people come to use the name, the meaning of which is anchored in the initial bestowal of the name in the baptism. For Putnam (1975), on the other hand, the meaning of mass and kind terms is determined by the environment in which you find yourself, such that two internal duplicates placed in environments which differ solely in the fact that the clear potable liquid in one is h2o while in the other it is xyz mean different things by their uses of 'water'. Yet a third view, associated perhaps most famously with some work by Timothy Williamson (1994), the meaning of a word is determined by its overall patterns of use, a pattern involving many many people, most of whom are and will always be strangers to each other.

The core feature of each of these views is that meanings are determined by something other than the speaker. But now that causes a problem. Say you're a would-be conceptual engineer. You realize that a given word has a bad meaning, and you want to change it. How do you go about it? In the simplified story I give above, the word will either be determined by causal chains, or the environment, or the uses of a whole linguistic community. But the would-be engineer won't have control over causal chains, environments, or the dispositions of the whole linguistic community.

If this is so, then conceptual engineering is a somewhat melancholy project. Cappelen thinks it is, but he thinks that that's just the way it goes with normative theorizing in general.

When concentrating on natural languages like English, Cappelen's thought strikes me as pretty compelling, and in particular it strikes me that the externalist metasemantics of natural language is not something over which would-be engineers have control. But earlier we noted that the metasemantics of social media are not quite the metasemantics of natural languages. The question I'll ask in the next section is whether these different metasemantics help us respond to Cappelen's worry. I'll suggest that they do.

4.1 Engineering Social Media

The clue is kind of in the name, and in particular in the word 'engineering'. Natural languages are not engineered: they arise spontaneously, and their metasemantics are determined not by any one person, but by a combination of causal chains, natural features, and use. But--and this is the important part--social media languages are, literally, engineered. They are the product of software engineers, an individual or more likely small team of individuals who have great control over the metasemantics of the language of twitter. If they wanted to, they could change what we do in tweeting tomorrow. They could, for example, remove reception information from tweets. If you buy my argument that that information is part of the semantic content of tweets, that would change the semantic content of millions of communications each and every day.

Say that we decided, on reflection, that including reception information was a bad feature of social media languages. I won't make that case, but we can note that it's prima facie plausible: there's a risk that one's evaluation of a tweet will be swayed, not by the text that it contains, but by the information about how it's been received. This is especially dangerous when we consider the malicious uses to which social media can be put. If being retweeted confers legitimacy on a tweet, then given one can make easily make new twitter identities and have them retweet it, one can artificially make a tweet appear more legitimate than it is, and spread falsehoods.

So we might conclude that it's good that reception information be removed. But that is eminently doable. We don't need to mess with causal chains, or the environment, or sprawling use of millions of strangers. Instead, we need to change some lines of code, and that small, localized change, will propagate semantic consequences throughout the whole system.

The metasemantics of twitter communication are changeable. What this means is that the conceptual engineering project, even if its prospects for natural language are somewhat grim, still has the potential to apply to the realm of online communication.

Admittedly, this will be cold comfort to many proposed instances of conceptual engineering. We could maybe distinguish two sorts of conceptual engineering: lexical engineering and structural engineering. In the former case, we are concerned with particular words. Kevin Scharp, for example, in *Replacing Truth*, is concerned with the word 'true'; Sally Haslanger is concerned with 'woman' and 'race' (see the papers collected in her 2012); Clark and Chalmers are concerned with 'knowledge'.

But a language--the thing we're trying to fix--is more than its lexicon. A language includes, for example, structural syntactic rules as for example the fact that in English one needs to provide a verb at least one syntactic argument even if it takes no semantic arguments: thus, in English one must say 'it rains', rather than 'rains' (as one can say in many languages), despite the fact that nothing is doing the raining.

Roughly, the sort of meaning we've been considering--reception information--is something similar to structural meaning. Just as it's obligatory that all verbs have syntactic arguments in English, so it's obligatory that all tweets have reception information, regardless of whether or not it's important or desired.

The point is, certain structural features of a given internet language are very apt for engineering by, to repeat my pun, engineers, and so Cappelen's metasemantic worry, although well-posed for lexically engineering natural languages, arguably has considerably less bite for the structural engineering of social media languages.

This is, I think, a happy conclusion, for many. For conceptual engineers, it gives them a domain to work. Moreover--maybe this is overly optimistic but--it gives them a chance to have a real world difference. The big tech companies at least pay lip service to the thought that they want to improve our online experiences, to foster better communication and to reduce trolling and misleading. I've argued that that's a conceptual engineering project, because part of what these companies are in the business of doing, even if they don't realize it, is creating new languages, and so, if they want help, they should appeal to conceptual engineers. But even for non-conceptual engineers who are merely worried about the world and about how online communication might be for the worse, it should give solace, because it provides a way to make the problem intelligible, and therefore to suggest a space of solutions. The problem is no longer the vague one with which we started of how to fix online communication but the considerably more precise one of how to engineer the metasemantics of social media platforms to remove negative features and add positive ones, and that's a question that a heap of philosophers will be able to help answer.

Bibliography

Ayer, A. J. (1936). *Language, Truth and Logic*. London: V. Gollancz.

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

Cappelen, Herman (forthcoming). Conceptual Engineering: The Master Argument. In Herman Cappelen, David Plunkett & Alexis Burgess (eds.), *Conceptual Engineering and Conceptual Ethics*. Oxford: Oxford University Press.

Cappelen, Herman & Lepore, Ernie (1997). On an Alleged Connection Between Indirect Speech and the Theory of Meaning. *Mind and Language* 12 (3-4):278–296.

Chomsky, Noam (1959). A review of B. F. Skinner's Verbal Behavior. *Language* 35 (1):26--58.

Clark, Andy & Chalmers, David J. (1998). The extended mind. *Analysis* 58 (1):7-19.

Egan, Andy (2009). Billboards, bombs and shotgun weddings. *Synthese* 166 (2):251-279.

Frege, Gottlob (1884). *Der Grundlagen Der Arithmetik*.

Haslanger, Sally (2012). *Resisting Reality: Social Construction and Social Critique*. Oxford University Press.

Heim, Irene & Kratzer, Angelika (1998). *Semantics in Generative Grammar*. Blackwell.

Heim, Irene,& Von Fintel, Kai (ms). *Intensional Semantics*.

Kripke, Saul (1980). *Naming and Necessity*. Harvard University Press.

Plunkett, David & Sundell, Timothy (2013). Disagreement and the Semantics of Normative and Evaluative Terms. *Philosophers' Imprint* 13.

Putnam, Hillary (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science* 7:131-193.

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

Williamson, Timothy (1994). *Vagueness*. Routledge.