On Retweeting

Abstract: A small but growing literature in philosophy is devoted to the understanding of a seemingly new communicative action that came with the internet, and with Twitter in particular: the retweet. The spur for this literature is a kind of puzzle in public discourse: on the one hand, there is a tendency to hold people responsible for their retweets, and to blame them for retweeting material considered offensive or otherwise inappropriate. On the other hand, there is a widely shared, if not universally-recognized feeling that, as the well-known disclaimer has it, "A retweet is not an endorsement." But if a retweet is not an endorsement, what is it? And what is wrong with retweeting offensive or misleading tweets? Here, we put forward the view that bare, uncommented retweets are best understood as lacking any sort of default illocutionary force, in contrast to many other types of speech acts. And whereas with most speech acts, it is massively difficult to influence the norms governing that type of act, this may not be the case with retweeting. With retweeting, we can consider ways that Twitter's interface and code might be altered so as to revise the act of retweeting; likewise, we can consider the likely impacts of such revisions on the norms surrounding retweeting. This raises a pair of interesting questions: (i) what should we want retweeting to be like and (ii) how can we make progress in that direction?

During the town hall, the President... tried to separate himself from his recent retweet of a conspiracy theory from an account linked to QAnon, which baselessly claimed that former Vice President Joe Biden orchestrated to have Seal Team Six killed to cover up the fake death of al Qaeda founder Osama bin Laden.

"I know nothing about it," Trump claimed. "That was a retweet -- that was an opinion of somebody. And that was a retweet. I'll put it out there. People can decide for themselves."

But Guthrie responded: "I don't get that. You're the President. You're not like someone's crazy uncle who can retweet whatever.

-- CNN Politics, Reported by Maegan Vazquez, 'Trump again refuses to denounce QAnon'

1. Introduction

A small but growing literature in philosophy is devoted to the understanding of a seemingly new communicative action that came into being with the development of the internet, and Twitter in particular: the retweet. The spur for this literature is a puzzle in public discourse. On the one hand, there is a tendency to hold people responsible for their retweets, and to blame them for retweeting material considered offensive or otherwise inappropriate. On the other hand, there is a widely shared, if not universally-recognized, feeling that, as the well-known disclaimer has it, "A retweet is not an endorsement." But if a retweet is not an endorsement, what is it? And what is wrong with retweeting offensive or misleading tweets? What sort of responsibility do people have for their retweets?

The first of these questions—*what is a retweet, if not an endorsement?*—asks for retweeting to be placed under some broader category of speech action. One way to do this, as the question acknowledges, is by classifying retweeting as a form of endorsement. To do so is to see retweeting as similar or analogous to, for instance, saying "I agree" in response to someone else's assertion, signing one's name to a petition, or giving a thumbs-up sign when shown a photograph someone is suggesting to use for some purpose. Effectively, the claim that retweeting *is* a broad form of endorsement proposes to understand retweeting as having a default illocutionary force: *prima facie* at least, retweets are 'assentives', in the terminology of Bach and Harnish (1979, 43).

Below we will review some compelling reasons against categorizing retweeting in this way, and pursue answers to the more general framing of this question—*what type of speech action is retweeting?*—by considering which other kinds of actions are communicatively similar to retweeting. We won't limit our answer to this question to existing categories of speech actions (e.g., what one might find in Bach and Harnish 1979), and will consider the possibility (raised, for example, by Marsili 2020) that retweets are an entirely new kind of speech act. Ultimately, however, we will suggest that retweeting has no default force and resists subsumption under illocutionary categories, old or new.

This conclusion might seem disappointing with respect to the further aim of answering the normative questions above—*what is wrong with retweeting offensive or misleading tweets?*—and, more generally—*what responsibility do people bear for their retweets?* If we could place retweeting in a broader speech category with other actions whose associated responsibilities are more widely established, then that would be one way to make the case for a particular set of associated responsibilities. The disclaimer "A retweet is not an endorsement" presumes that endorsement brings with it certain responsibilities, and stipulates that these should not be applied to retweeting. But if we can identify no illocutionary category in which retweeting, by default, belongs, then judgments about responsibility associated with such categories do not shed light on the question of responsibility for one's retweets.

There is, however, another way of approaching the effort to categorize retweeting—one which opens the door to a rather different way of proceeding. On this alternative approach, instead of aiming to describe what kind of speech act retweeting is in the here and now, we aim to describe what kind of speech act retweeting ought to be. Call this the 'engineering' project with respect to retweeting. While philosophers are used to distinguishing between the non-ideal question of what our norms regarding X are at present and the ideal question of what they ought to be like, it's not a distinction that has been applied specifically to the case of speech acts. However, we believe that this distinction may prove especially important in the case of novel speech acts like retweeting.

Of course, answers to all these questions—*what sort of speech act is retweeting? What are the norms surrounding it? And what ought those norms to be?*—are clearly interrelated. If the best answer to the first of these questions is that retweeting is an endorsement, then the answer to the second should be fairly clear—and, supposing that our present norms of endorsement are in good working order, perhaps the third as well. As we'll see below, there is something distinctly interesting for philosophers about the phenomenon of retweeting: retweeting is a communicative act that we have explicitly engineered for ourselves, by way of the large and influential social media company, Twitter. Most of the speech acts with which we are familiar did not arise like this; rather, assertions, requests, inquisitives, etc. presumably all developed out of repeated interactions between individuals, with the understanding of each of those individuals partially determining how subsequent uses of a particular grammatical construction or innovation were likely to be interpreted. The phenomenon of retweeting was created largely by engineers who designed the platform, with a bit of code. Unlike our social evolution, that code can be tweaked in a relatively straightforward way—via the intervention of Twitter's engineers.

What this means, practically, is that we can seriously contemplate substantial, top-down revision of the way that the act of retweeting functions. This would be almost inconceivable when it comes to more established speech acts like asserting or commanding.[1]

Here's the plan for what comes next. In section 2, we will consider three proposals (including an earlier one of our own) that seek to categorize bare retweeting within extant illocutionary types. As we'll explain, we take all of these proposals to be mistaken: retweeting is more basic than these proposals would have it, lacking any kind of default illocutionary force. In sections 3 and 4, we clarify what we take to be the interesting philosophical project of speech act *engineering*, a project of asking not what speech acts we presently have, but what speech acts we ought to work to

[1] That said, we do believe that the new forms of interaction introduced by way of social media—the new ways of asserting or requesting—may eventually change, even quite drastically or fundamentally, even well-established speech acts like asserting or commanding. We will have to defer consideration of this issue to another occasion, however.

promote because they better serve our legitimate purposes (communicative, epistemic, moral, etc.). In section 3, we explore proposals that would aim to engineer retweeting so that it fits within a well established illocutionary type. In section 4, we consider how we might engineer retweeting, supposing that our descriptive account of retweeting is and should remain correct. Finally, section 5 concludes.

2. Retweeting as a basic act

The question 'What sort of speech act is retweeting?' can be seen as a question about the illocutionary force of retweeting, in the traditional Austinian sense. The question might also be put as 'What does one *do* in retweeting a tweet?' It should be agreed, though, that the fact that an action is a (bare) retweet[2] does not in itself entail that the action is any particular sort of illocutionary act. This is analogous to the fact that an utterance of a declarative sentence like "Snow is white," can constitute a variety of different illocutionary acts in different conversational contexts. A bare retweet, like an utterance of "Snow is white," is simply a basic communicative tool within a particular communicative system. In the case of "Snow is white," the system is the spoken version of the natural language English. In the case of the bare retweet, the system is Twitter. (Obviously, these two communicative systems interact substantially.)

Even if it is clear that an act's being a retweet does not entail its having any particular illocutionary force, one might suppose that an act's being a retweet would dispose it toward having a certain illocutionary force—that there will be a default illocutionary force attached to a bare retweet, even if that default can be overridden. This would be in the same way that an act's being an utterance of "Snow is white." seems to dispose it to be an assertion, or an act's being an utterance of "Close the door!" seems to dispose it to be a command. We can ask whether retweeting has a default, or dispositional, illocutionary force, in the way that utterances of sentences of various types arguably do.

For reasons we are about to set out, we think it is difficult to identify any such default force for retweeting. Retweeting a particular kind of tweet lacks the sort of literal meaning or content that makes an act of uttering a declarative sentence well-suited to be an assertion, or an act of uttering an imperative sentence well-suited to be a command. The fact that it is difficult to identify any such default force for retweeting is problematic, because it makes it hard to tell what speech act people are performing when they retweet. It may even make it difficult to perform distinctive speech acts by retweeting, especially in cases where retweeters do not share rich informational backgrounds with their audiences. The good news, as

[2] In what follows we will often use simply "retweet" instead of "bare retweet". We will always mean bare retweets unless otherwise specified.

mentioned above, is that retweeting—unlike uttering a sentence of English—is an act whose features can be programmed and re-programmed without the tacit cooperation of an entire linguistic community. Retweeting can be "engineered" so as to make acts of retweeting better-suited to being acts with a particular illocutionary force (or forces), and less well-suited to being acts with other sorts of illocutionary forces. (This engineering will be the topic of section 3 and 4.) In the remainder of this section, we will make the case that retweeting is not like other speech acts that seem to have a default illocutionary force.

We start by considering three models of the default force for retweeting: endorsement, reporting, and indication.[3] These models yield the following analogies:

A user, U, retweeting a tweet, T, by original poster, OP, is like:

(1) U uttering, "I endorse this," where the demonstrative "this" refers to T. (endorsement)
(2) U uttering, "OP tweeted this," where the demonstrative "this" refers to T. (reporting)
(3) U uttering, "Look at this," where the demonstrative "this" refers to T. (indicating = directing attention)

Both Arielli (2018) and Marsili (2020) argue that the popular disclaimer, "a retweet is not an endorsement" suggests that retweeting is not like (1). If retweeting were something like (1), it would seem contradictory to issue such a disclaimer. But, intuitively at least, this disclaimer does not seem contradictory. Further, if one is questioned about why one retweeted a certain tweet, it does not seem contradictory to reply, "I just thought people should see it—I don't endorse it." (Marsili 2020, 10464) Similar points could be made about any model of retweeting that treats it as a way of expressing approval for the original tweet. If retweeting were the same as, "I approve of this," or "I like this," or "Hurray for this!", the popular disclaimer would instead seem nonsensical.[4] So retweeting is not akin to endorsing in any of these ways.

Regarding (2), both Arielli and Marsili point out that to report what someone says, either by direct quotation or indirect discourse, is to make an assertion, which is true or false depending on whether that person spoke as reported, and which is sincere or insincere depending on whether the reporter believes that that person

[3] In choosing these alternatives, we follow Arielli (2018) and Marsili (2020).
[4] In his discussion, Marsili focuses on a narrow notion of endorsement, according to which endorsing a tweet is expressing agreement with the content of the tweet. He argues that if retweeting meant something like, "I agree with the content of this [original tweet]", then it would seem contradictory to claim that one does not agree with a tweet that one retweeted; but it does not seem contradictory to do this. He also argues that if retweeting meant something like this, then most tweets would not be appropriate to retweet, since most tweets are not of the right form to be agreed with. Here we are suggesting that the first type of argument works equally well against analogies with various ways of expressing endorsement or approval more generally.

spoke as reported. In contrast to this, a retweet is not true or false depending on whether the OP tweeted the tweet. A retweet *shows* that the OP tweeted the tweet, and is not evaluable as true or false. If the OP did not tweet the tweet that a putative retweet appears to retweet, then the putative retweet is not a retweet at all (but rather some sort of fake). Likewise, if it makes sense to posit sincerity conditions for retweeting in itself at all (which we doubt), it seems most plausible that a retweet is sincere if and only if the particular act of retweeting is a genuine expression of the retweeter's beliefs or feelings concerning the content or context of the original tweet and/or subsequent uptake. It surely does not turn on whether they believe that the OP tweeted the tweet. So retweeting is not like reporting on someone's tweet by stating that they tweeted it.

The observation that a retweet *shows* a tweet, rather than reports on it, suggests that the right model for retweeting is indication or *attention-direction*. This is a model we previously suggested for online re-sharing in general, retweeting included (Pepp, Michaelson, and Sterken 2019), which also finds some favor with Arielli and Marsili. It likens retweeting to pointing communicatively at the original tweet, or, as suggested in (3), issuing the instruction, "Look at this!" There is much to commend this model over the others: it does not make denial of endorsement contradictory, and it does not require retweets to be truth-apt or sincerity-apt in ways they clearly are not. Still, we are now of the opinion that it is not quite right.

Both Arielli and Marsili argue that to retweet is not *merely* to indicate or direct attention to the original tweet, since a retweeter also reproduces the original tweet. This reproduction creates a new opportunity for people to see the tweet, thereby amplifying it or expanding its reach in a way that simple indication does not. We agree with this. But we think there is a further problem with treating retweeting on the model of (3). This problem, which is similar to the problem pointed out for (1), suggests that retweeting does not have the default force of directing attention to the original tweet.

To see the problem, consider the following example. Suppose that @Nora, a Twitter user who is a huge fan of Roblox, retweets a post of, "Roblox rules!" by her favorite Roblox YouTuber star, @it'sakeila. Suppose further that (i) all of @Nora's Twitter followers are also enthusiastic followers of @it'sakeila and so have seen the post, and (ii) that @Nora's followers, being such dedicated fans who keep in touch about all aspects of @it'sakeila, all know that all of them have seen the post and know that they know that they've seen the post, and so on. That is, suppose that the tweet by @it'sakeila is common ground amongst @Nora's followers. Now imagine that @Nora's mother, who struggles to understand the ways of these young internet denizens, observes @Nora's retweet and asks her daughter, "Why did you retweet that if you've all seen it already and know that each other have all seen it already?" @Nora rolls her eyes and replies (in the exasperated tone reserved for parents), "I wasn't telling them to *look* at it! I was just showing that I'm into it!" If retweeting had

a default meaning along the lines of "Look at this," @Nora's reply would seem contradictory. But it does not. For comparison, imagine instead that @Nora is on the playground with her friends, and suddenly grabs a sand pail and says, "Look at this!" If her mother asks her, "Why did you tell them to look at that pail?", it would be nonsensical for @Nora to reply, "I didn't tell them to look at it, I was just showing that I'm into it!". Given this, it seems incorrect to assimilate retweeting to a locutionary act like uttering, "Look at this," or even a non-linguistic act of directing attention by pointing.

One concern regarding this argument stems from the observation that one can say, "Look at this/that)," without intending to instruct someone to look at the thing referred to. For instance, it is common to react to a beautiful sight that one's companions have already attended to and remarked upon by saying, "Yeah, wow, look at that," or the like. Given this, retweeting might well have a default meaning similar to uttering, "Look at this/that," and still be well suited to be used non-directively, simply to bond with others.[5] This objection is more compelling if it is generally true that locutions whose default meaning is attention-directing are also easily understood as fostering bonding over already-attended items. If this is a robust, cross-linguistic pattern—something we are simply unsure about—then the @Nora case fits nicely into an account on which retweeting is like saying, "Look at this," or pointing at something. If it is not, then the way in which retweeting seems to pattern with the English sentences, "Look at this" or "look at that" may be more coincidental.

It is also worth noting that physically pointing at things already jointly attended to and remarked upon seems generally out of place. Suppose A, B and C are out walking. A points at the sunset and says, "Look at that, how beautiful!" B says, "Oh, wow, what a great sunset!" It would be natural enough for C to chime in by saying, "Yeah, look at that." But it would be downright strange for her to join in by silently pointing at the sunset, and even slightly strange for her to point at it while saying, "Yeah, look at that." So, to the extent that we think that retweeting isn't merely just like saying "Look at this," but also involves something like pointing to the relevant tweet, @Nora's response here would be unanticipated by the present model.

It seems fair to say that the use of the retweet, as in the @Nora case, calls (3) into question as a model of retweeting, but is not decisive against it. Even if (3) is not the right model, it may be feasible to treat retweeting as a more basic form of indicating; a form of indicating that does not carry the default prescriptive, attention-directing force of locutionary acts like uttering, "Look at this." For instance, Marsili (2020, 10472-74) analyzes retweeting as overtly showing or making manifest the original tweet. He uses this base analysis to build a general relevance-theoretic

[5] Thanks to Rachel Fraser and audience members at the Jowett Society for pressing an objection along these lines.

account of the communicative value of retweets. Marsili's account is consistent with our suggestion here, that retweeting does not have a default meaning or illocutionary force, since he understands indicating as something more basic than any type of illocution, something which figures into illocutionary acts via its interaction with patterns of relevance to particular listeners. In contrast, in our earlier work we took indicating to be a thicker, more prescriptive kind of activity—more in line with the model represented by (3).

At this point, we are going to leave aside this descriptive question regarding what the norms of retweeting are. For, as interesting a question as this is, we are more interested in the question of what the norms of retweeting could be—and what we might be able to do to push them in that direction.

3. Engineering the retweet

As noted at the start, philosophers are sometimes interested not just, or not even primarily, in offering descriptively adequate accounts of this or that phenomenon, but in offering normative accounts of how those phenomena *ought to be*. Such projects are perhaps more familiar in ethics—focused, as it is, on how human beings ought to behave—and political philosophy—focused, as it is, on how humans ought to behave, together, at various levels of idealization. More recently one strand of this project has come once again to prominence in the philosophy of language and mind, under the rubric of *conceptual engineering*.[6,7]

According to those engaged in the project of conceptual engineering, we should not only ask ourselves what characterizes a certain concept currently in circulation, WOMAN for instance, and why, but also what other revised or replacement concept(s) might serve us better than our present one, along various epistemic, ethical, and political dimensions. Disagreements persist about both how to weigh these various factors and just how far from our current practice it is legitimate to stray while retaining a claim to be modifying an extant concept as opposed to introducing a new one, but we can safely ignore these for present purposes.

With respect to retweeting, it is not really the concept that is at issue. Rather, it is the background social norms and practices and, perhaps most saliently, the speech act itself. So, for instance, one might take it that the indication model of retweeting offers an accurate account of our present norms of retweeting, but that the endorsement model would nonetheless be *better* for us in various ways if it were true. Rini (2017, E-55) recommends something along these lines: we should *aim* to make the endorsement model true, even though it very likely isn't now.

[6] For a relevant overview, see Eklund (2021) and Cappelen & Plunkett (2020).
[7] It's worth noting that this isn't the first time that conceptual engineering has played a prominent role in the philosophy of language and mind. The logical positivists, for instance, can be viewed as having been engaged in such a project. Likewise, in the Anglo-American tradition, see Ambrose (1952).

In contrast to the project of conceptual engineering, where we try to grapple with which concepts might serve us better or worse along various dimensions, the present project is one of *speech act engineering*, a project where we ask what kinds of speech acts and accompanying social norms will serve us better or worse along various dimensions. Like conceptual engineering, speech act engineering isn't a new philosophical endeavor. Rather, it is a part of a broader long-standing project of *normative engineering*, a project of asking which sort of norms—be they political, ethical, epistemic, legal, aesthetic, etc.—we *ought to* adopt in order to promote goods of various kinds, or perhaps just the good in general. For example, *The Republic* contains an extended inquiry not into what the norms surrounding family life were like in ancient Athens, but rather into what such norms *ought to be like* in order to help promote a more ideal political arrangement. Likewise, political philosophers tend not to be interested in what our norms of distribution are now, but rather which norms would serve to maximally promote justice or overall well-being under one or another set of idealizations.

The interesting thing about speech act engineering in the present context is that we are dealing with a very recent, and explicitly constructed, set of speech acts—something which holds out the possibility that these speech acts could be practically *re-engineered* in ways difficult to conceive of when it comes to e.g. assertions or commands. Twitter tweaks its algorithm all the time, and the notions of 'liking' and 'retweeting' are hardly immutable; these are acts which were introduced into the platform at a particular time, whose feel and effects can be and have been altered via the adjustment of a few (thousand, or tens or hundreds of thousands of) lines of code, or via the introduction of external moderation of the platform. In other words, this is a place where philosophers have—at least in principle—a relatively clear picture of some of the mechanisms by which normative changes might be affected. So let us ask: given what we know about Twitter today, how might we hope to re-engineer the retweet?

Rini offers what we take to be some helpful initial suggestions here, starting with two concrete claims : (i) we ought to re-engineer retweeting so as to inhibit the spread of fake news on Twitter; and (ii) coming to a point where retweeting is seen as endorsement and carries accountability norms in accord with this will help us to realize this aim. Here is what Rini has to say on the matter:

> If we firmly established the norm that social media sharers are understood as conveying testimonial endorsement, then people would be less likely to share unverified stories, to avoid later being held responsible for errors. Alternatively, if we firmly established the norm that social media shares (without further comment) communicate no testimonial endorsement whatsoever, then people would be less likely to come to believe fake news on the basis of their friends' transmissions. (E-55)

Rini discounts the second alternative for being unrealistic, and advocates pursuit of the first. She suggests that the establishment of an endorsement norm would allow us to deploy our ordinary ways of holding people to account for what they say—by applying to retweets the forms of criticism ordinarily suited to testimony. That, she reasons, should help to prevent the proliferation of fake news on the platform.

The question we want to ask is: how might we hope to operationalize Rini's proposal? That is, what might Twitter, Twitter users collectively or the government regulatory apparatus do to establish a testimonial norm for retweeting?

One very high-level suggestion would just be to advocate for a collective shift of consciousness; obviously, whether a bare retweet constitutes an endorsement depends in part on the norms surrounding bare retweeting, and those in turn depend partly on our attitudes towards tweeting and retweeting. Rini suggests one way that Twitter might help to promote the adoption of such a collective shift of consciousness: Twitter might provide infrastructure to track the testimonial reliability and reputation of particular users (2018: E-57). This suggestion effectively attempts to mirror, within Twitter, the way that our pre-social media norms of endorsement function and are maintained: we track, both individually and collectively, the reliability of our interlocutors when they make assertions and endorse others' claims. If Twitter were to mimic this offline behavior, this line of thought runs, then we can imagine that, over time, we might come to treat (at least many) retweets as speech acts which are subject to evaluations in terms of reliability.

Another natural option would be to suggest a ban on pure retweeting. All retweeting could require the retweeter to add a coherent comment. One would need an algorithm to prevent people from circumventing this ban by simply mashing the keyboard or writing gibberish—but, presumably, at least some of that could be prevented. One worry, however, is that this would invite a sort of cat-and-mouse approach by sub-groups looking to spread stories without taking full responsibility for their spread. So one might instead think that the thing to do is to require that all tweets be valenced—with something like a thumbs-up, thumbs-down, laughter, and perhaps a range of other reactions. Or perhaps one should be forced to commit to a take on how accurate the thing is that one is sharing. So a bare retweet would come with a tag like 'very accurate' or 'unsure of the accuracy'. Granted, this sort of re-engineering would likely be put to some comical effects when it comes to retweeting cute cat photos (very accurate!), but if our goal is to slow the spread of fake news then perhaps that's a price we should be willing to pay.

How helpful should we expect any of these suggestions to be, if implemented? We're rather skeptical that either would be of much help—though we are happy to grant that this is ultimately an empirical question, and one that could fairly readily

be tested.[8] Our qualms stem mostly from the observation that, in our current, highly-partisan atmosphere, fake news seems to have become largely tailored to those with a certain package of social/political views and a tendency towards skepticism towards those who they identify as of a different political identity. So criticism, either personal or institutional, of one's endorsement of some fake content seems to us likely to have the opposite effect of what we might have hoped: rather than prompting the targets of that content to tamp down their credence in it, it may well reinforce their credence and prompt them to decrease their trust in the 'other side', be that some individuals or rather Twitter itself. Again, controlled tests of Twitter-like platforms would likely tell us much more about the conditions under which tweaks like these might attain their desired effects and when they might not.

None of this is meant to constitute a direct argument against Rini's proposal, or the more general idea that we ought to try and make it the case that retweets are subject to the same norms as endorsements.[9] Rather, our aim here is just to point out that projects in speech act engineering, like other projects in normative engineering, are subject to an *implementation challenge*. The interesting thing about these projects when it comes to speech acts that arise in highly engineered environments like Twitter, is that we could plausibly answer questions like 'Would doing X reduce the spread of Y?' with some degree of accuracy without having to undertake massive social change. None of this is to claim that the answers to questions like this one are likely to be either easy or intuitive—just that they are somewhat more tractable than many that we as philosophers have considered in the past.

4. A broader view of the speech act engineering project

So far, we have followed Rini in accepting that the aim of re-engineering the retweet is to reduce the spread of fake news and disinformation. And while we certainly agree that this ought to be *one* important aim of this project, we strongly suspect that there will be others as well.[10] So let us step back and ask: what other problems are there that might be addressed by re-engineering the retweet? What benefits are there to be had? And what norms would serve to minimize these problems while

[8] Some data might come from studying Twitter's current initiative to reduce misinformation online in the run-up to and aftermath of the US Presidential election, which involves, among other measures, changing the retweet function so that whenever someone retweets a tweet, they are automatically prompted to make it a commented retweet rather than a pure retweet, by adding their own commentary. This is not as extreme a change as the one suggested above, which would *force* commented retweeting (it is still possible to do a pure retweet by leaving the comment field blank). But it would be interesting to study whether this change leads to less pure retweeting, and whether it leads to less retweeting in general of misinformation.

[9] Indeed, since we take Rini's proposal to be introduced primarily for the purpose of illustration, we hardly think that Rini would disagree with anything we have claimed above (E-58).

[10] Though we disagree on other important issues (cf. Pepp, Michaelson & Sterken 2019b and Habgoode-Coote 2020), we take this to be a point well made in Habgood-Coote (2019).

maximizing the relevant benefits? In other words, what is the weighted overall aim of social media (Twitter, in this case), and what sorts of norms will best serve to promote that weighted overall aim?

To ask this set of questions is, effectively, to acknowledge another aspect of the scope and difficulty of the project of engineering the retweet. To carry out that project, we will need to identify and weigh a range of different aims we might have for the use/practice of Twitter. We also need to determine which norms for retweeting are most conducive to each of these various aims and how those norms interact with each other. A comprehensive treatment of this kind is too much to attempt in this chapter. Indeed, as we pointed out in the previous section, determining which norms for retweeting are conducive to any given aim (for example, whether an endorsement norm is conducive to the aim of limiting fake news) will involve substantial empirical work. So if we are confined to our armchairs, we must in any event content ourselves with working hypotheses about what sorts of speech norms should be engineered.

Here we will start with a still more modest plan. In section 4.1, we will consider a speech act engineering proposal that seems to be favored by some very different aims from those that might favor Rini's. On this proposal, the norm for retweeting to be engineered fits with what we earlier called the *indication* account. We rejected this as a *descriptive* proposal about the illocutionary force of retweeting. By considering it now as an *engineering* proposal, we will explore one set of aims it might promote, which probably do not include the aim of limiting the spread of fake news. Still, these aims are worthwhile. This exercise will serve to emphasize that a broad view of the proper aims for social media is needed in order to carry out the speech act engineering project successfully.[11]

We will close, in section 4.2, by suggesting that in light of this requirement, there is much to recommend *not* working to attach specific norms to retweeting as such. Instead, it might behoove us to work toward increased appreciation that retweeting itself is just a use of a flexible communicative tool (like uttering "Snow is white"), which can itself be put to various further purposes. Using this tool without any evident purpose is thus (at least) an odd thing to do (like uttering "Snow is white" without any evident purpose). As Savannah Guthrie pointed out to President Trump in the epigraph, it is something you might expect and tolerate from a 'crazy uncle', but which one ought not to tolerate from experts, politicians, and the like.

More importantly, though, this outlook suggests that the norms for retweeting we should work toward will be local norms, tailored to different social networks, topics, and social positions of the retweeters. Accordingly, this will mean that responsibility

[11] Note: as above, the hypotheses we will suggest about the relation between norms for retweeting and aims for Twitter are just that, and not worth much until tested empirically. The point of the discussion in 4.1 is to illustrate a plausible scenario in which different, worthwhile aims for Twitter could favor different norms for retweeting.

for one's retweets is a highly contextualized affair, depending not only on the fact of having performed the basic act of retweeting, but on which illocutionary acts are thereby performed in a given situation, and what the significance of these acts is given the topic of discussion and public roles of the participants.

4.1. Aims promoted by indication-type norms

Twitter, we contend, can be productively viewed as a kind of social system, or game even, within which various actions are available (cf. Nguyen 2021). Certain aspects of this game are arguably rather negative, but thinking about Twitter as a collective, participatory activity provides a different lens on its currently prevailing norms, as well as on the ways in which those norms should be changed.

It also prompts the question of what we all are collectively doing by being active on Twitter, which might not merely be the conjunction of what each of us individually are doing. (And we can ask the same about other social media platforms, and about social media in general.) One idea is that we are collectively structuring available perspectives on the internet. We need to create these perspectives because the internet is such a vast space of information. One way to interact with this sea of information is to try to find some specific information within it, for instance by using search engines or knowledge of relevant websites to tell you where to look. But another way to interact with the internet is simply to go online and see what is there, without looking for anything in particular. The difference is analogous to the difference between looking for a black cat in your visual field as opposed to just taking in your visual field without looking for anything in particular. But the internet does not have a ready analog for the visual field—a limited space of information over which one may (voluntarily or involuntarily) direct one's attention. In order to go online and see what is there, each of us needs some way of creating an online visual field. Twitter, and social media more generally, is a common tool for this. We rely on other people—those we (mostly) choose to be connected to, and by extension those they are connected to—to structure our attention online. Collectively, then, one thing we are doing with social media is creating perspectives, or viewpoints, on the internet.

It is not so far-fetched to suppose that creating these perspectives on the internet is also an *aim* we should have for Twitter. The internet offers its users the opportunity to gain a broader, less filtered awareness of political and cultural events, opinions and perspectives, works of art, and different ways of life than was ever possible before. But some narrowing and filtering is required for users to be able to realize this opportunity. Twitter (among other social media platforms) has the potential to let users provide this narrowing and filtering for one another, without the need to appoint any authorities or rely solely on algorithms to do the job.

It is plausible (but, again, ultimately an empirical matter) that establishing an indication-type norm for retweeting would promote this aim. If retweeting were universally seen as an act of making manifest one's intention to make salient the retweeted content by replicating it (in the distinctive way enabled by Twitter), then the norm for retweeting might be something like this: retweet only if you wish to promote the visibility and accessibility of the original tweet within the Twitter perspectives of your audience. Establishment of this norm would, at least to some degree, free Twitter users from worries about accuracy and appropriateness that might otherwise stop some of them from retweeting tweets they find worthy of visibility. Establishment of this norm would transform retweeting into a quick and easy 'vote' to keep a piece of content present in people's online visual fields. Establishment of this norm would also free Twitter users from any felt demand to comment on retweeted content. To satisfy this norm, it is enough to want the tweet to be visible and accessible, and one's reasons for wanting that are another matter. Indeed, commenting or otherwise making explicit one's attitude to the retweeted content could interfere with the aim of structuring perspectives on the internet, since it could have the effect of making the comment, rather than the retweeted content, the primary or 'at-issue' content of the retweet.[12]

It is worth emphasizing that the aim of creating perspectives on the internet is also the aim of creating what we might call 'action-points'. In retweeting, one does not just promote the visibility of the original tweet, one also provides an opportunity to interact with it. This interaction can take many different forms, from further pure retweeting, to liking, to replying, to commented retweeting. Thus, if the present hypothesis is correct, establishing an indication-type norm of retweeting would not only promote the aim of providing people with a manageable view of the internet, but also the aim of providing them with a manageable way to interact with the internet and further develop the available viewpoints and action-points.

The goal of this section has not been to argue that we ought to promote an indication-type norm. Promotion of such a norm could have significant downsides, such as the legitimation of attempts to spread disinformation along the lines of Trump's answer to Guthrie's questioning (in the epigraph). What we wish to take away from the foregoing is that promotion of such a norm could also have upsides, and could help to fulfill some legitimate aims that we might have for Twitter and other forms of social media. The broader lesson is that it seems inevitable that different legitimate aims will favor different norms for retweeting. Thus, the project

[12] Indeed, this effect is, in a certain way, built into the structure of commented retweets (i.e. 'quote retweets'), since they create a new post that gets attention, acquiring likes and further retweets instead of the original. Because of this, many artists who promote their artwork in part by accumulating such statistics and monitoring comments on their own tweeted work view quote retweeting as damaging and rude. See, for example: https://www.theverge.com/2020/10/21/21527101/twitter-retweet-changes-artists-quote-retweet-qrt https://twitter.com/shiroganejpg/status/1113599683035897857?lang=en

of engineering norms for retweeting requires not only determining which norms are favored by certain aims, but also determining which aims should be given the most weight in driving the project, and how different aims should be balanced against one another when they inevitably come into conflict.

4.2. Normativizing the basic act approach

With the complexity of the project of speech act engineering in better view, it seems clear that one thing we might want is for retweeting to be governed by different norms in different situations. For instance, we might want the retweeting of certain types of content to be governed by different norms than the retweeting of other types. Perhaps it would be best for retweeting of (apparent) news articles to have an endorsement or assertion norm, while retweeting of opinion pieces or funny stories should have an indication-type norm. But from here it is only a small step to notice that we will probably want the norms to depend on the identity and status of the retweeter, as well. The President of the United States retweeting someone's opinion gives that opinion a prominence that a little known private individual's retweet could never offer. It would probably be good if Guthrie were right that the norms of retweeting are thereby violated when the President retweets something that he merely wishes to "put out there" but not endorse. Nor will the need for subtler norms stop with a dependence on the nature of the retweeted content, or the position and status of the retweeter. We will want different norms for those retweeting a content friendly to their known views than for those retweeting a content unfriendly to those views; different norms for those retweeting on a private Twitter than on a public one; different norms for retweeting a public figure's tweet versus a little known person's tweet. And so on.

In short, we are going to want the norms for retweeting to be as variable and situation-dependent as the norms for uttering sentences. This suggests that the quest to choose an illocutionary category as the target one for retweeting, and to try to influence people's behavior so that retweeting in general will be viewed as part of that category, is probably misguided.

What, then, should the normative engineering project for retweeting look like? It should focus, we think, on instituting specific, local interventions or nudges designed to encourage the development of desirable norms for retweeting in specific kinds of circumstances. As we have been at pains to emphasize, the questions of which nudges or other structural changes to platforms would encourage the development of which norms, and of which norms would lead to the achievement of different aims and goals for Twitter, are largely empirical questions. Nonetheless,

we'll briefly outline some concrete suggestions in the spirit of illustrating the shape of the speech act engineering project we favor.[13]

We offer this outline in a similar spirit to Rini's proposal: as an illustration of the kind of thing we think it would be productive for social media companies to try and test. In contrast to Rini, however, our proposal targets specific communicative circumstances with the aim of promoting certain norms for retweeting in those circumstances—as opposed to retweeting in general.

So, for the sake of illustration, let's consider the aim of reducing the spread of fake news on Twitter while safeguarding the attention-structuring role of retweeting that we discussed in the previous section. To do this, we should develop interventions that are targeted at a certain type of tweet: those that are apt to be treated as *news.* This type of tweet has many sub-types. It will include tweets of article links from news websites, or websites that appear to be news websites. It will include tweets from journalists and media companies directly reporting what they present as news. It might include tweets from state and industry actors that simultaneously *make* and *report* news, as when Donald Trump tweeted on election night 2020 "I will be making a statement tonight. A big WIN!" With this tweet, Trump made news by announcing that he would make a statement and simultaneously reported that news. (He also falsely implied that it was already clear he would win the election.) Announcement tweets by people and institutions whose announcements constitute news of wide interest have a status similar to press releases or press briefings. Interventions could be targeted at retweets of any of these types of tweets via algorithms developed to recognize such retweets.

Having singled out news retweets, we would then want to categorize the retweeters according to influence. Various dimensions of influence could and should be recognized. Let's just consider three: number of Twitter followers, status as a public figure, and status as a government official or office.

Number of followers is an easy metric to track. Public figure status is more difficult, but Twitter could use legal precedents from defamation actions to label certain users as public figures. Status as a government official or office should be fairly straightforward to identify. With these classifications in hand, different interventions could be targeted at public figures as compared with relative unknowns, at government officials as compared with private citizens, and at users meeting a series of different thresholds for follower numbers. For instance, users with over 1000 followers (only 2.12% of Twitter users) might be required to have a label show up on all news tweets and retweets, either "[Twitter name] attests to the accuracy of this piece of news" or "[Twitter name] does not attest to the accuracy of

[13] It is worth pointing out that the need for empirical input into this project, to understand both the link between interventions and norm development and the link between norm development and outcomes, means that it is unavoidably an interdisciplinary project, requiring collaboration with social and information scientists.

this piece of news". Users with over 20,000 followers (only 0.06% of Twitter users) and public figures might not be allowed to tweet or retweet news without the first label on it, attesting to its accuracy. Finally, government officials and offices could be barred from tweeting or retweeting news at all, with the exception of news-making tweets of the kind just described. This would prevent them from using the power and prominence of their office to influence journalism as it is now practiced. Twitter and social media more broadly is now an important part of standard journalistic distribution practice, so preventing government actors from influencing the spread of news on those platforms could be seen as a way of promoting a continued free press. The categorization and identification of tweeters/retweeters in this way might also help to address the issue of fake accounts and their role in the spread of fake news and structuring of public attention/discourse.[14]

A system like this would leave most Twitter users free to tweet and retweet news as they like, "just to put it out there", and to help structure each other's perspectives on the internet. But it would be made clearer that those with significant influence are expected to commit one way or the other on accuracy, and those with even more influence must stand behind the accuracy of news to which they provide a platform. The only part of this system that is sure to achieve its aims is the barring of government officials from tweeting and retweeting news (at least from their official accounts). For influential tweeters of other stripes, it could happen that the required attestations would be viewed as pro forma and mostly be ignored by users, hence failing to create a norm of endorsement or a norm of accuracy-attestation around such retweets. But it is also possible that at least weaker norms would be created: perhaps it would come to be expected that one would choose the accuracy attestation only if one had fairly strong credence in the accuracy of the original tweet, for instance. And if such norms did develop, they might also trickle down, perhaps in yet weaker forms, to less influential users in their retweeting practices. This is the sense in which a measure like the one suggested would be (largely) a nudge rather than a rule.

5. Conclusion

Above, we surveyed extant accounts of retweeting's default illocutionary category —and argued for their descriptive inadequacy. Instead, we suggested, the best descriptive account of bare retweets is that they are a basic type of act in the Twitter environment. As basic communicative acts, without a default illocutionary force, there are few, if any, generalizations to be drawn about what makes bare retweets appropriate. We take that to be an acceptable outcome, however, since it shifts attention to what strikes us as the more interesting question: what makes bare retweeting acceptable in this or that kind of context?

[14] None of this is to say that we think Twitter is at all likely to take up our suggestions.

We have suggested that there may, at present, be no one clear answer to this. Still, there is a closely related question of great interest: how ought we to re-engineer this communicative environment so that users are effectively required, or at least collectively nudged, to accept certain norms of conduct in certain contexts? As a first step, we suggested that we should better differentiate certain sorts of communicative contexts that will regularly arise on Twitter. In particular, we proposed to differentiate retweets based on *who* is doing the retweeting and *what* they are retweeting, with different or stricter standards applied to better-known individuals, or those occupying certain social roles or public offices, when they retweet items of broad epistemic importance. This, it seems to us, is likely to be a good starting point for improving our collective epistemic situation—assuming that we continue to rely on platforms like Twitter as an important aspect of our epistemic engagement with the world. Undoubtedly, more than this will need to be done, including implementing some of the further distinctions we mentioned in the previous section. And here, in contrast with many other parts of philosophy, we take there to be real scope for philosophers to put forward concrete suggestions that might well help to improve our future social media landscape.

References

Ambrose, A. (1952). Linguistic approaches to philosophical problems. *The Journal of Philosophy*, *49*(9), 289-301.

Arielli, E. (2018). Sharing as speech act. *Versus*, *47*(2), 243-258.

Austin, J. L. (1975). *How to Do Things with Words* (Vol. 88). Oxford university press.

Bach, K., & Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. MIT Press.

Cappelen, H. (2020). Assertion: A defective theoretical category. In *The Oxford Handbook of Assertion*. Oxford University Press.

Cappelen, H., & Plunkett, D. (2020). Introduction: A guided tour of conceptual engineering and conceptual ethics. In *Conceptual engineering and Conceptual Ethics*. Oxford University Press.

Eklund, M. Conceptual engineering in philosophy. In *The Routledge Handbook of Social and Political Philosophy of Language*. Routledge.

Habgood-Coote, J. (2019). Stop talking about fake news!. *Inquiry*, *62*(9-10), 1033-1065.

Habgood-Coote, J. (2020). Fake news, conceptual engineering, and linguistic resistance: reply to Pepp, Michaelson and Sterken, and Brown. Inquiry, 1-29.

Marsili, N. (2020). Retweeting: its linguistic and epistemic value. Synthese.

Nguyen, T. (2021). How Twitter Gamifies Communication. In *Applied Communication*. Oxford University Press.

Owens, D. (2012). Shaping the Normative Landscape. Oxford University Press.

Pepp, J., Michaelson, E., & Sterken, R. K. (2019a). What's new about fake news. J. Ethics & Soc. Phil., 16, 67.

Pepp, J., Michaelson, E., & Sterken, R.K. (2019b). Why we should keep talking about fake news. Inquiry, 1-17.

Rini, R. (2017). Fake news and partisan epistemology. Kennedy Institute of Ethics Journal, 27(2), E-43.