



This is a repository copy of *Retweeting: its linguistic and epistemic value*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/161612/>

Version: Accepted Version

Article:

Marsili, N. orcid.org/0000-0001-7853-7359 (2021) *Retweeting: its linguistic and epistemic value*. *Synthese*, 198 (11). pp. 10457-10483. ISSN 0039-7857

<https://doi.org/10.1007/s11229-020-02731-y>

This is a post-peer-review, pre-copyedit version of an article published in *Synthese*. The final authenticated version is available online at: <http://dx.doi.org/10.1007/s11229-020-02731-y>.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Retweeting: its Linguistic and Epistemic Value

Neri Marsili¹

Abstract

This paper analyses the communicative and epistemic value of retweeting (and more generally of reposting content on social media). Against a naïve view, it argues that retweets are not acts of endorsement, motivating this diagnosis with linguistic data. Retweeting is instead modelled as a peculiar form of quotation, in which the reported content is indicated rather than reproduced. A relevance-theoretic account of the communicative import of retweeting is then developed, to spell out the complex mechanisms by which retweets achieve their communicative goals. The last section outlines the epistemic threats posed by the increasing prevalence of retweeting on social media, linking them to the low reputational, cognitive and practical costs linked to this emerging form of communication.

Keywords Online communication; Retweeting; Speech Act Theory; Testimony; Fake News.

1 Retweeting: a novel communicative act

1.1 Microblogging, information flow, and fake news

“Repeat a lie often enough and it becomes the truth”
Popular phrase, often attributed to Joseph Goebbels

A systematic study of how communication works in contemporary societies cannot prescind from understanding the phenomenon of microblogging: the online broadcasting of short texts, images and links on social media. As of 2018, about two in three people living in Western countries use social media regularly to communicate (Nic, Levy, and Nielsen 2018). The emergence of microblogging has had an impact not only on the way people communicate, but also on the way that information circulates in our society. Social media are a source of news and information for two thirds (68%) of American adults (Shearer and Gottfried 2017).

Contact: nerimarsili@ub.edu

¹ Faculty of Philosophy, University of Barcelona, Carrer de Montalegre 6, 08001, Barcelona, Spain

More generally, about one in four people (23%) now uses social media as their *main* gateway to news, with young adults (18-34) being more likely to access news on social media than on any other source (Nic, Levy, and Nielsen 2018, figs. Q10, Q10a).

As social media become a central hub for accessing and discussing daily news, there are growing concerns that this transition could have negative effects. There is evidence that like-minded users on social media get stuck in “echo chambers” or “filter bubbles”, where they are insulated from perspectives contrary to their own (Sunstein 2001; Pariser 2011). Furthermore, studies show that social media platforms are the main conduits for fake news sites (Lazer et al. 2018). During the 2016 US elections, popular fake news stories outperformed the most popular mainstream news stories in terms of visibility on these platforms (Silverman 2016), and the results of the elections were significantly affected by the visibility that social media gave to false and misleading information about both candidates (Allcott and Gentzkow 2017).

One of the most prominent ways in which news stories (including fake news stories, and rumours) achieve visibility and go viral online is by *retweeting cascades* (Vosoughi, Roy, and Aral 2018): one user posts an assertion or a link to a news article, and others then propagate it by retweeting it, in a ‘cascade’ that can reach millions of users. Retweets (and more generally reposts on social media) thus play an important role in the spread of information and misinformation online. Users (often bots, cf. Shao et al. 2018) who retweet veridical or mendacious content contribute to determining how much exposure said content achieves, and how many people it reaches.

In retweeting or sharing a post, users perform *communicative actions*; to learn what retweeting is, then, is in great part to understand which kind of communicative action is performed by issuing a retweet. Despite the significance of retweeting and sharing in the circulation of information and misinformation in contemporary societies, however, there has been little or no systematic academic discussion of their communicative and epistemic value. The aim of this paper is to fill this gap.

1.2 Speech acts on social media and speech acts of social media

It is uncontroversial that microblogging involves performing speech acts with illocutionary force, in the traditional sense (e.g. Austin 1962; Alston 1964; Searle 1969). By publishing a ‘post’ on Twitter or Facebook, a user can *assert* a proposition, *ask* a question, or issue a *request* – in other words, perform illocutionary acts. The work of speech act theorists on microblogging has so far focused on how ‘ordinary’ speech acts (e.g. promising, apologising, requesting, ...) are performed on online platforms, and how their performances differ from their non-virtual counterparts (Page 2014; Dayter 2014; Wikström 2014). Scholarly work on microblogging also includes quantitative research aimed at measuring which illocutionary acts are most frequently performed on these platforms (Vosoughi and Roy 2016; Carr, Schrock, and Dauterman 2012).

It is hard to deny that social media platforms also allow users to perform communicative actions that they would *not* be able to perform in ordinary offline communication: for instance, to *like* or *react* to a post, to *share* it or *retweet* it. Not much scholarly work, however, has attempted to analyse communicative acts that

can be performed *exclusively* on these platforms (but cf. Labinaz and Sbisà 2020). In a number of cases, the communicative value of these actions is admittedly straightforward: the ‘Angry’ Reaction² on Facebook (😡) is a way to express anger, irritation or discontent; the ‘Wow’ Reaction (😲) expresses surprise, and so forth. Other times, however, the communicative value of these online speech acts is less clear: by clicking on the heart-shaped button to ‘like’ a tweet, a user may express approval, surprise, amusement, agreement, or merely give away an intention to give the target post more visibility. Liking a tweet surely has some communicative value, but its exact communicative import is vaguer and more undefined than the one conveyed by the emoji-based ‘Reactions’ of Facebook. Similarly, retweeting is a speech act that we can only perform online, and whose communicative import is not entirely obvious. The aim of this paper is to offer a systematic account of its communicative value, and to model the epistemic expectations that it engenders.

In this paper, the term ‘retweeting’ is used to refer, quite straightforwardly, to the act of reposting a tweet, performed by means of the built-in ‘Retweet’ function provided by the Twitter platform. Other social media platforms provide equivalent features (such as the ‘share’ function on Facebook) that allow users to share existing posts to their followers. This paper will focus on retweeting for ease of discussion, but its analysis of retweets equally applies (with minor contextual differences) to equivalent functions on other social media platforms.

Here is how I will proceed. The next section introduces two hypotheses about the communicative value of retweets: the *endorsement* view and the *quotation* view. Discussing these views will allow me to present and discuss some key linguistic data points that a theory of retweeting should accommodate, and to identify some potential misunderstandings that should be avoided. I will suggest that the quotation view offers a more promising framework to understand retweets. But the analogy with quotation only brings us so far: in section 3 I will show that retweeting differs from standard quotation in some important ways. Section 4 outlines a positive account of the communicative value of retweets, and section 5 identifies the principal epistemic threats posed by the increasing prevalence of retweeting on social media, linking them to the low reputational, cognitive and practical costs of this new form of communication.

² Readers unfamiliar with the Facebook ‘Reaction’ feature may consult Liz Stinson, ‘[Facebook Reactions, the Totally Redesigned Like Button, is Here](#)’, *Wired*, 24/2/2016.

2 Two hypotheses about the communicative value of retweets

2.1 Setting the scene: some methodological remarks

(T1)



Consider the above retweet, issued by Donald Trump on the 26th of October 2019³. What has Donald Trump done in retweeting T1? That is, what is the illocutionary force of the speech act performed by the president of the United States, if any? A natural answer is that Trump has somehow *expressed agreement* with the statement posted by Lynn Thomas: he endorsed the claim that democrats are the only ones who are interfering in the US elections. Alternatively, it could be suggested that in retweeting T1 Trump has simply pointed his followers' attention to the fact that someone (Lynn Thomas) has claimed that democrats are the only ones who are interfering in the US elections, without taking an explicit stance as to whether he agrees or not with this assessment of the current political situation.

More generally, we begin our analysis by noting that there are at least two ways of conceiving the communicative import of a retweet, which lie at the opposite ends of a spectrum: at one end, retweets are conceived as endorsements of the original tweet; at the opposite end, retweets are treated as mere speech reports that take no explicit stance about the value of the original tweet. I will consider these two accounts in turn, to see what each has to say about the meaning of retweeting⁴.

³ Here is a quick guide for the reader who is not familiar with Twitter. T1 is a snapshot of a retweet. The first line indicates the username (Donald J. Trump) of the person who retweeted the message. The original tweet follows below. On the first line of the original tweet we see the username of the person who originally posted it (LYNN THOMAS), and that user's Twitter handle (@LYNNTHO06607841). The text of the original tweet follows. The last line displays the number of comments, retweets and likes that the original tweet received. Clicking on the four icons at the bottom, you can respectively add a comment, retweet the tweet, like it, or send a message to the original poster. Note that this image is merely a reconstruction of the original retweet: just hours after it was posted, @LYNNTHO06607841 was suspended for violation of Twitter's terms of service, and all its tweets were erased from the platform.

⁴ Let me stress since the outset that by selecting these two hypotheses I do not wish to suggest that there are no other options to model the communicative value of retweets, nor that these are the most theoretically promising accounts that one could come up with. Considering these two accounts is

Before proceeding, however, I would like to outline some important methodological challenges one faces in analysing retweets, and how I plan to address them. First, unlike traditional speech acts (like promising or commanding), retweeting can only be performed by means of an artificial tool (a social media platform) designed for commercial purposes, whose features set it apart from traditional (verbal and written) communication. This limits the extent to which traditional approaches to the study of communication apply to its analysis. Because of this, at times we will inevitably meet some methodological challenges (such as the absence of a clear theoretical framework for analysing the notion of representation in digital environments, cf. §3.2). Since the purpose of this paper is not to provide a general framework for studying online communication (but rather to analyse retweeting specifically), we will have at times to rest content with brief, tentative answers to the bigger methodological questions that arise in this new field of inquiry, leaving some of them open for future scholarship to address.

Second, and relatedly, retweeting is a linguistic phenomenon that emerged in relatively recent times. We are concerned with conventions and norms that are still in the process of being established, and that may change significantly in the near future (cf. Labinaz and Sbisà, forth:8, Rini 2017:5). What is true today of what retweets communicate may not be true in a few years. That being said, there are observations to be made about what retweets *can* and *cannot* communicate, given their current structural features (for instance, given the fact that they can convey content that is non-sentential, cf. §2.2, or given that they necessarily attribute content to *someone*, cf. §3.3). Consequently, this paper will advance two different kinds of claims: unqualified, general claims about how the structural features of the Twitter platform constrain the communicative value of retweets, and *contingent* claims about the value that retweets have currently acquired, within these constraints. Which kind of claim is advanced in any given case will be either clear from the context, or explicitly specified.

Finally, it is important to highlight that the relevant objects of this inquiry are *uncommented, non-ironic* retweets. Comments are themselves speech acts, and

instrumental to introduce and discuss some important features of retweets (and clear some potential misconceptions), laying the groundwork for the rest of the paper, in which I present my positive proposal. Beyond these motivations, note that one reason to discuss the *endorsement view* specifically (and its shortcomings, cf. §2.2) is that this view is often advanced or presupposed in public and legal discussion about retweets (cf. §2.2 and §5.1, and Metaxas et al. 2015:§2.2.3).

A referee suggested that I could offer a brief overview of some of the alternative accounts that this paper (due to its space limitations) cannot directly discuss. For ease of exposition, we may divide them into two families, based on whether they come closer to the *endorsement view* or to the *quotation view*. Among the former, we may mention the view that retweets *express trust* in the reliability of the information contained in the tweet, or in its *source* (Metaxas and Mustafaraj 2013:254); or that they signify endorsement *of the person who issued the tweet*, rather than endorsement of its content. Among the latter, we have the view that retweets merely express a *desire to promote* the original tweet; that they constitute an *invitation to attend to its content*; that they signify a “vote for the quality, novelty or timeliness of a piece of information” that it conveys (Van Liere 2010); or, more simply, that they *communicate that one finds the tweet interesting*. Note that most of the objections that will be considered in §2.2 also put some pressure on the former family of views, and that the view developed in §3-4 is able to accommodate most of the communicative functions highlighted by the latter family of views. So, although this paper cannot engage directly with each conceivable view, reasons in favour and against each can be found throughout it. For a more complete conceptual map of different conceptions of the meaning of retweets, cf. Boyd, Golder, and Lotan (2010, 6), and Metaxas and TTRT (2017).

speech acts (e.g. two assertions) can cancel or amend each other (Caponetto 2018), just like (2) cancels the assertoric value of (1):

- (1) I love you
 (2) Just kidding: I only wanted to see you blush!

Similarly, the addition of a disclaimer (e.g. ‘this fake news is hilarious’), or an explicit endorsement (‘I totally agree with this’) to a retweet is in itself a speech act, and as such it can substantially alter the meaning of the retweet to which it is attached. To analyse the communicative value of retweet, it is thus best to limit our attention to uncommented retweets.

Irony can also drastically affect the meaning of a speech act. Most of the 120.000 users who retweeted the infamous (and meaningless) ‘covfefe’ tweet (T2) clearly did not mean to express their agreement with Trump’s nonsensical statement, but rather bafflement and hilarity:

(T2)



Although I will argue that the communicative function of retweets is not to express agreement, these nonserious uses do not challenge the view that retweets are endorsements, because irony is often antiphrastic. To identify the communicative value of a retweet is to identify the value that a *serious* retweet has, *before* it is modified by other speech acts (such as those performed in a commented retweet)

The endorsement hypothesis

Let us start by considering the first of the two accounts of retweeting sketched in the previous section. According to what we may call the ‘endorsement hypothesis’, retweeting simply amount to *endorsing* the original tweet. On this view, in T1 Donald Trump has endorsed or espoused⁵ Lynn Thomas’s claim, expressing his agreement with what is said in the tweet. Retweeting is thus something like uttering (4) or (5) in response to (3):

- (3) Democrats are the only ones who are interfering in the US elections
 (4) I agree that they are

⁵ Labinaz & Sbisà (forth) use this terminology, in a footnote where they observe that uncommented reposts seem to receive the default reading of an *espousal*.

(5) True that!

In uttering (4) or (5), one *agrees with*, *expresses approval of*, or *endorses* the assertion (3). On the proposed view, the force of a retweet (its illocutionary import) is analogous to the force of (4) and (5): it is to express agreement with the original tweet, to endorse it. In agreeing with (3) by uttering (4) or (5), the speaker makes a conversational move whose communicative effect is broadly equivalent to asserting (3) directly (cf. Alston 2002:38). The difference between asserting and endorsing is that the latter speech act is parasitic to someone else's assertion: in endorsing, you assert indirectly, *by agreeing* with someone else's claim. To sum up, the endorsement view holds that in retweeting a tweet you express your agreement with it, thereby reaffirming its content⁶.

A first and simple linguistic observation in support of the endorsement account is that we typically interpret retweets as endorsements, and that we also typically intend them to be interpreted as such. Upon being presented with a retweet, and lacking any contextual reason *against* interpreting a retweet as an endorsement, by default we take the retweeter to be agreeing with the content of the original tweet. For instance, upon being presented with T1, a competent user will typically take Trump to be expressing agreement with the tweet posted by Lynn Thomas.

However, the expectation that a retweet expresses agreement is *defeasible*, and in this it differs from explicit endorsements like (4) or (5). Retweeters can void such presumption simply by specifying that their retweet was not meant as an endorsement. For instance, upon being questioned about why he retweeted T1, Trump may respond (perhaps implausibly, but without contradiction) that he merely wanted to invite discussion on the topic, or that he only wanted to inform his followers that a particular user (Lynn Thomas) had claimed that democrats are the only ones who are interfering in the US elections, although he disagrees with such an assessment of the political situation. More generally, users can successfully cancel the presumption that specific retweets are to be interpreted as acts of endorsement: for future reference, let us call this feature of retweets *cancellability*.

Politicians and public figures often exploit cancellability to avoid being held accountable for what they retweet (more on this in §5.1): when a retweet causes controversy in the public arena, it is not uncommon for their issuers to dodge criticisms by denying that they meant to express agreement with the content they shared, even when it is rather obvious from the context that that was their original intention⁷. A good account of the communicative value of retweets should be able to explain what goes on in these cases, and clarify why these denials are available to retweeters, even when their initial communicative intention is clear from the context.

⁶ Let me stress that my adoption of the term 'endorsement' (which is motivated by its use in online discussion about the communicative value of retweets) should not be taken to suggest that any attitude that could be described as 'endorsing' would satisfy the mooted view. Rather, the intended meaning of the word is the one stipulated in the text: for a retweeter to 'endorse' a tweet is for them to express *agreement* with the content of a tweet.

⁷ For a discussion of recent examples of such attempts (both of the plausible and of the implausible sort), see Jennifer Mercieca, "[The insidious strategy behind Donald Trump's retweets](#)", *Maclean's*, 30/11/2017, and "[Donald Trump's favorite excuse? It was just a 'retweet'](#)", *Indiana Express*, 23/2/2016

A related phenomenon is the increasing number of journalists, reporters and media producers that feature the disclaimer “A retweet is not an endorsement” in their Twitter profile description (Metaxas et al. 2015, sec. 2.2.3; Rini 2017, 5–6). While these users are not able to postulate (à la Humpty Dumpty) *any* meaning for their retweets, they seem to be able to have a say on how they want their retweets *not* to be interpreted. Specifically, they seem to be able to successfully determine that their retweets do not have the force of endorsements. Call this feature of retweets *negotiability*.

Both negotiability and deniability point in two different directions. On the one hand, they suggest that there is a default interpretation of retweets as endorsements. It would be pointless to add a disclaimer like “a RT is not an endorsement” (or to insist that a specific retweet was not meant as such) if retweets were only rarely (or never) interpreted as endorsements. On the other hand, the fact that speakers are able to successfully establish that their retweets are not to be interpreted as endorsements is incompatible with the hypothesis that this is the communicative import (*i.e.* explicit meaning, force) of retweeting. If the value of retweeting is to agree with a target proposition (or set thereof), as the endorsement view has it, then denials of this kind should sound absurd or paradoxical. This becomes clear if we consider what it would mean for a speaker to deny that the point of a particular instance of agreeing was expressing convergence of belief (6) (the would-be equivalent of *cancellability*), or to try to negotiate the value of agreeing in general (7) (the would-be equivalent of *negotiability*):

(6) # When I said that I agree that peas are legumes, I did not mean to suggest that I believe that peas are legumes

(7) # Whenever I say that I agree with you, I do not mean that I share your opinion

Neither (6) or (7) are reasonable things to say: both utterances are clearly infelicitous. While (6) is quasi-Moorean (the speaker denies having the psychological attitude expressed by the first half of the sentence), (7) is purely Humptydumptyian (it postulates a new meaning for the word ‘agree’). If retweeting really amounted to explicitly agreeing or reasserting the target sentence, it should similarly be absurd to deny that in retweeting you meant to agree with the target tweet, and Humptydumptyian to claim that ‘a retweet is not an endorsement’. Since these conversational patterns are instead familiar, common, and perceived as felicitous, cancellability and negotiability speak against the proposed analogy with endorsements.

Could a weaker version of the endorsement view accommodate this data? Perhaps it could be argued that while retweets are not acts of endorsement, the *default implicature* (as defined by Grice 1989) stemming from a retweet is that the retweeter endorses the original tweet. This would neatly explain why the endorsement expressed by (some) retweets is both cancellable and negotiable, and why there is nothing absurd in denying that your retweets are meant as endorsements: if agreement is merely implied, the implication that the speaker agrees with the original poster (henceforth, OP) can easily be taken back (cancelled) or denied. I find this view *prima facie* appealing, and I will defend one that delivers similar predictions in §4. But there are two reasons why the *default-implicature*

version of the endorsement view faces compelling problems as an account of the *illocutionary force* of retweets.

The first is that while the default-implicature view tells us something about what retweets usually imply, it still leaves open the question of what the literal force of a retweet is – what is the primary, direct illocutionary act that one performs in retweeting. To determine what an illocutionary act can imply is not yet to determine the literal force of that act. For instance, an assertion may imply a request (“You are standing on my foot”), a threat (“I have a knife and I know how to use it”), or other speech acts, but knowing these facts does not in itself help us determine what an assertion is, or its communicative value. Since the default-implicature view only tells us what retweets typically imply, it is at most a piece of the explanation of the communicative value of retweets. It still leaves open our primary question, namely determining the *literal* force of retweets.

The second and more crucial reason why retweets cannot be treated as endorsements (be they direct or indirect) is that tweets are not necessarily something that one *can* endorse – so that in many cases endorsement is not even the *default implicature* of retweets. We have initially limited our attention to retweets of assertions, but one can retweet speech acts other than assertions. To see this, consider the following examples:

(T3)



(T4)



Neither T3 nor T4 are strictly speaking assertions: T3 is an invitation (or a request), T4 is a collection of different speech acts: a promise, an assertion, and an exclamation. The endorsement view would have that in retweeting T3 and T4, you *reassert* or *endorse* the claim made in the original tweet. But a retweet of T3 is typically interpreted as a request to sign the petition itself, and T4 simply does not seem the sort of linguistic item that one can endorse.

Perhaps the endorsement view just needs to be *refined* once again: in a loose sense of the term ‘endorse’, to endorse an invitation is to make that invitation yourself, and to *endorse* a class of speech acts is to express agreement with the general idea that they express (although it may be quite obscure how we are meant to identify ‘the idea’ that a collection of sentences expresses, especially in cases like T4). But attempts to amend the endorsement view in this direction are bound to fail, as there are many retweets whose content cannot be endorsed in any relevant sense of the term.

(T5)



Consider T5. Clearly, no competent twitter user would claim that by retweeting the photo of a cute kitten Margaret Pittman has endorsed (directly or via implicature) the cute kitten, or that she has affirmed something about the kitten: photos are not something you can reassert or agree with. Importantly, kitten pictures are not an exception: tweets can contain videos, images, pictures, drawings, animations, and all sort of non-sentential content– and these are all things that one cannot simply ‘endorse’.

The endorsement account is therefore bound to be inadequate⁸. At most, it may apply to a suitably defined subset of retweets – say, retweets of *assertoric tweets*. But tweets contain material that cannot easily be parsed this way: they may feature assertions, sets of assertions, assertions mixed with other speech acts, graphs that contain pseudo-assertoric content, and comic strips and videos in which several and even contradictory illocutionary acts are performed. We should be sceptical of the possibility of individuating non-arbitrary boundaries in this field. Any account of retweeting that only applies to a suitably defined subset would not only have a limited explanatory value (limited to *that chosen subset*): it would also run the risk of being both arbitrary (in defining the subset) and ad-hoc (in ruling out that which does not belong to the subset). Quite importantly, the alternative hypothesis that we are about to consider easily avoids this sort of difficulties.

As a final point, it should be noted that the default-implicature view would find it hard to accommodate this data even if we were to understand the notion of ‘default’ in a looser sense, for instance as the claim that *typically* retweets are used to communicate or imply endorsement. This is because ‘endorsable’ content

⁸ Rini (2017:12) has argued that we should *aim* to adopt the endorsement account – aim to make it become the conventional value of retweets. I agree that this would be epistemically desirable, as it would allow people to really hold each other accountable for what they retweet, limiting the spread of misinformation (cf. §5). Against these optimistic remarks, however, it follows from the present analysis that we cannot converge on such a convention, due to the very fabric of the Twitter platform. That is, unless Twitter significantly changes how retweets work, retweets *cannot* be endorsements, as an important subset of retweeted content is not the sort of thing that one can endorse.

constitutes a minute portion of what is posted on social media platforms. Assertive speech acts account for only about one third of sentential posts (*i.e.* posts containing sentences) on Twitter, and sentential posts in turn account for a rather limited subset of the overall content that circulates on the platform (Vosoughi and Roy 2016). Statistically, tweets that *can* be endorsed are rather uncommon: it follows that it is false that *typically* retweets can be used to communicate or imply endorsement, against the revised hypothesis.

The quotation hypothesis

According to the ‘quotation hypothesis’, retweeting is nothing as committal as endorsing the original tweet – it is simply to show that someone has tweeted something, making the content of that tweet available to your followers. Understood in these terms, retweeting is a mere form of speech report, broadly equivalent to quoting someone in offline speech – like using (9) or (10) to report what Jenny says in (8):

- (8) The restaurant opens at 5 pm
- (9) Jenny said: “The restaurant opens at 5 pm”
- (10) Jenny said that the trattoria opens at 5

Like endorsements, speech reports are higher-order speech acts, as they take another speech act as content. But unlike endorsements, reports do not typically take a stance towards the lower-order speech act: they merely reproduce it, in a forceless fashion akin to mere mentioning (Davidson 1964; Saka 1998). This is evident in (9) and (10): here the speaker is not taking an explicit stance on whether (she agrees that) the restaurant opens at 5 pm – she is merely reporting that Jenny said so.

This view can easily accommodate the linguistic data considered so far. The most compelling objection against the endorsement view (that tweets are not necessarily something one can endorse) hardly represents a challenge for the quotation view. Quotation can clearly target speech acts other than assertions (and collections thereof), and there is nothing controversial in saying that by retweeting T3, T4 or T5 I would be reporting what the OP⁹ has posted.

More generally, it is recognised that quoted content need not be sentential nor verbal (Searle 1969; Goodman 1974, 305–6; Fleischmann 1999; Rossholm 2004), and scholars are familiar with cases of non-verbal quotation, like (11):

- (11) So Jane tells him that she just had a baby, and he’s like [speaker makes a facial expression of shock]

It may seem, by contrast, that the quotation view may have problems explaining why retweets of *assertoric content* are often interpreted as endorsements. But speech reports *can* be used to endorse the reported speech act, either implicitly or explicitly. Explicit endorsement is expressed when the reported content is

⁹ Adopting a popular acronym, in what follows I will abbreviate “original poster” (the issuer of the original tweet) as OP.

accompanied by some expression that conventionally specifies that the speaker is endorsing (or denying, doubting, etc.) it. For instance, (12) expresses agreement to the target statement:

(12) As Jenny said, “The restaurant opens at 5 pm”

Unqualified (plain) quotation, by contrast can only convey the speaker’s attitude towards the target utterance rather indirectly. At most, utterances like (9) or (10) can *suggest* or *imply* that the speaker agrees with the original utterance. If we are trying to establish at which time it is best to go to the restaurant and I utter (10), not only I will be taken to have *claimed* that Jenny said that the restaurant opens at 5pm, but also to have *implied or suggested* that Jenny’s claim was informative, and that 5pm is indeed the time at which the restaurant opens. More generally, in reporting what someone says, a speaker can often imply that the reported content is true, or that they agree with it (cf. §4.2, §5.1).

This fits rather well with the picture delineated so far. Retweets do not convey explicit endorsements, and we have seen that the same is true of unqualified quotation. Furthermore, quotation at most conveys a *defeasible* presumption of agreement, a limitation that sits well with the *cancellability* and *negotiability* of retweets. Reporting that *p* is compatible with specifying that one does not agree with *p*, so that there is nothing absurd in (13) or (14) (which are the counterparts of (6) and (7), the absurd constructions that I derived from the endorsement view):

(13) Jenny said that the restaurant opens at 5pm, but I disagree

(14) When I report what other people have said, I don’t necessarily mean that I agree with their opinion

The linguistic data considered so far suggests that retweets *report* rather than *endorse* the content of the original tweets. But surely this dichotomy does not exhaust the range of possible illocutions that retweets can perform, and a more nuanced view can be developed. In the next section, I will attempt to outline in which respect retweets are similar to quotations, and in which respect they are different from them.

3 Retweeting as quoting, and beyond

3.1 Direct and indirect representation

On a standard approach to quotation, quotations are *metarepresentations* (Recanati 2000; Wilson 2012): that is, representations that take representations as contents. Scholars often classify different kinds of quotation depending on how the target content (the tweet, in our case) is represented. If we restrict our attention to reported speech, a standard distinction opposes *direct quotation* (or *oratio recta*, like (9)), to

indirect quotation (or *oratio obliqua*, like (10))¹⁰. Direct quotation involves directly reproducing the original speech act, whereas indirect quotation involves indirectly *describing* what was communicated by that speech act:

(9) Jenny said: “The restaurant opens at 5 pm”

(10) Jenny said that the trattoria opens at 5

Retweeting has more features in common with direct quotation than indirect quotation: it involves *reproducing* the original signal, rather than indirectly *describing* it. But there are important differences between direct quotation in ordinary language and in retweeting. First, retweets reproduce the target utterance (the tweet) in ways that are quite unlike standard quotation (§3.2). Second, the truth-conditional meaning and force of quotation are not assimilable to those of retweets (§3.3). Let us see these differences in turn.

3.2 Representation and ostension: retweeting as ‘quoting by indicating’

Let us consider how direct quotation, as opposed to retweeting, reproduces the original representation. A speaker who directly quotes another is neither expected nor able to produce a representation of the target utterance that is *perfectly identical* to the original one. In a verbal speech report like (9), I would not typically be expected to mimic prosody, accent, or the timber and colour of the Jenny’s voice. In written quotation (such as the opening quote of this article), similarly, one is not expected to reproduce the original handwriting style, or irrelevant features of typewritten text (such as size, font, colour and layout). Rather than plain re-tokening, direct quotation involves the production of a stimulus that is similar to the original one *in some relevant respect* (Wilson 2012, 141-2). Furthermore, speech reports may *misrepresent* their target contents, either intentionally (in which case the report is *insincere*) or unintentionally (in which case the report is mistaken).

Retweets, by contrast, are inevitably *identical* to the target tweets: they cannot modify the content of the original tweet¹¹. Retweets cannot *misrepresent* the content that they report; the retweet is identical to the original tweet not just in some relevant respect, but rather *in every respect*¹². We may say in this sense that a retweet

¹⁰ To be sure, direct and indirect quotation do not exhaust the ways in which we can classify quotation based on how the reported content is represented (see Cappelen and Lepore 1997; Recanati 2000), but considering this basic opposition is sufficient for our purposes.

¹¹ Note that before the ‘native retweet’ was introduced at the end of 2009, the standard way to retweet a message was to copy-paste it after the ‘RT’ followed by the twitter handle of the OP (e.g. RT @user: original tweet) (cf. Boyd, Golder, and Lotan 2010, 3). These ‘non-native’ retweets sometimes misrepresented the original tweet, and it was not uncommon for retweets to omit part of the original message, since the addition of the “RT @user” syntax to the original message often caused the retweet to exceed the word limit for a tweet, forcing the retweeter to shorten the original tweet in order to make it fit. I will ignore this complication, as this paper uses the term ‘retweet’ to only refer to native retweets (rather than text-based ones), as discussed in §1.2.

¹² Clearly, if one visualises a tweet in one browser (or device, etc.) and its retweet in a different browser (or device etc.), differences may arise that are due to the features of the software (or hardware) on which the (re)tweet is tokened. But these differences are clearly irrelevant and beyond

is *transparent*, in the familiar sense in which this term is employed in aesthetics (Walton 1984): we see the represented content *through* it, as if we were looking through a window or a telescope.

In fact, the analogy with windows could be pushed a bit further. Retweets do not only allow users to see the original tweet: they make it possible to *interact* with it, for instance by liking it. Liking an uncommented retweet affects the like count of the original tweet, not that of the retweet. One can thus interact with the original tweet (click on it, like it) *through* the retweet. We may say, in this sense, that retweeting is more similar to *indicating* than it is to quoting, as indicating does not involve re-producing a given stimulus, but rather making it available to someone by directing their attention to it (Clark and Gerrig 1990, 765). But one important difference is that, unlike indicating, retweeting creates a new point of access to the indicated content – a ‘digital window’ onto it.

So retweeting has features in common with both *quoting* and *indicating*. It is an ostensive act: it makes manifest an intention to make something manifest. Like indications, it directs the audience’s attention towards something; like quotation, the target of this act of ostension is a *representation* (another communicative stimulus).

To help visualise what retweeting has in common with both quoting and indicating, consider the following example. My friends and I are discussing plans for the evening. As we walk and chat, I point my finger to indicate a billboard that advertises the next movie showing in the cinema: I *indicate* a *representation*. What I achieve by pointing at the billboard is rather similar to what I would achieve by quoting its content, for instance by saying: “The billboard says that *La Dolce Vita* is screening at the *Odeon* tonight”. When directed at representations, acts of *indication* can perform communicative functions rather similar to *quoting*.

Retweeting is an act similar to my pointing at the billboard: it is an act of indication that takes a representation as content: an act of ‘quoting by indicating’. But there is an important caveat in the proposed analogy. While ‘quoting by indicating’ does not produce new tokens of the indicated content, retweeting does: once retweeted, the original tweet appears on the profiles of both the OP and the retweeter. The retweet opens up a ‘digital window’ onto the tweet which is itself a new representation of the tweet, in a sense in which my finger pointing at the billboard is not a new representation of the billboard. In other words, retweeting retains an important feature in common with quoting and not with indicating, namely the fact that it does create a new occurrence of the ostended content, by representing it¹³.

the point, because they arise in the reception and visualisation of the message (on a specific device), not in the production of it. To put the point differently, they are not differences *between the tweet and its retweet*, but rather differences in how each is displayed on someone’s device.

¹³ Admittedly, matters are more complicated. I am helping myself with intuitive notions (such as that of a ‘digital window’) to refer to complex phenomena, and I am accepting some controversial assumptions (such as the idea that a retweet is a *representation* of the original tweet, even if it allows users to directly interact with it). As I anticipated in §2.1, one of the limitations of this work is that it ventures into a communicative environment where the traditional toolbox of pragmatics and philosophy of language simply does not apply. Since the aim of this paper is not to refine such a toolbox, I will have to limit myself to this level of approximation, leaving many fascinating questions (about identity and representation in digital environments) aside. For our purposes, it is

We can conclude that retweeting involves a peculiar form of ostension, analogous to an act of ‘quoting by indicating’, and that differs from it in that it creates a new occurrence of the target representation. For a lack of better term, I will call this form of representation a *digital ostension*.

3.3 Attributive use and ostension

Quotation (unlike retweeting, I will argue) is chiefly a matter of attributing a representation to *someone*¹⁴. More specifically, in quoting you make an *assertion* whose content is this attribution, and whose truth-value is determined by the correctness of the attribution. For instance, (9) is an *assertion* that attributes the utterance of “The restaurant opens at 5 pm” to Jenny. If Jenny indeed said it, (9) is true; if she did not say it, (9) is false. The *sincerity* of quotation, similarly, depends on whether the speaker believes that the attribution is correct or incorrect – in the example, on whether the speaker believes that Jenny said “The restaurant opens at 5pm”.

Can the same be said of retweets? Since the attributive function is built-into the Twitter platform (the OP is automatically included in the retweet), retweets *always* have an attributive component. In retweeting someone’s tweet, you also communicate that *that user* has posted something (cf. fn2). But this communicative function is marginal in retweeting if compared to quotation, for a number of reasons.

First, retweets have a different syntax from direct quotation (but cf. fn10). A retweet is not an attributive statement of the form ‘X posted T’ that can be true or false depending on whether the OP has posted something. Since a retweet cannot be true or false depending on the correctness of this attribution, there seems to be no ground to contend that such attribution determines the literal truth-conditional meaning of a retweet.

Second, and relatedly, retweets do not advance *claims* – they are not *assertions* about what someone has posted. Arguably, you assert a proposition *p* only if you are liable to be correct or incorrect depending on whether *p* is true or false (Green 2017). But retweets are not liable to be correct or incorrect in this sense. This is both due to the difference in syntax just discussed, and to the fact that retweets cannot be incorrect in their attributions, as explained in the previous section.

Third, users often lack an attributive *intention* in retweeting, and *interpret* each other as not having it. More often than not, the identity of the OP is rather irrelevant, for instance because the OP is not the author of the original post. Consider example T5 again: in all likelihood, @SoCuteKittens is not the author of most of the photos they publish. When users retweet a kitten photo from @SoCuteKittens, their primary intention is not to communicate that it was @SoCuteKittens who posted the photo: they merely want to make the original tweet (the kitten picture) available to their followers. Since an *attributive* retweet (that merely requires pressing a

sufficient to have established that retweeting has some important features in common with indicating that are absent in standard quotation, along the lines that I have sketched.

¹⁴ There are non-standard quotations that are non-attributive. For instance, neither (i) nor (ii) attributes the target expression to anyone in particular:

- (i) “Chitemmuort” is an expression in Neapolitan
- (ii) “Birds of a feather flock together”.

button) is less costly than a *non-attributive* new tweet (that would require downloading the kitten photo, reuploading it and posting it), people often choose to retweet despite having no intention to *attribute* a post to someone.

In sum, although retweets are necessarily attributive, attribution in retweeting lacks the central communicative role that it has in direct quotation: it does not contribute to truth-conditional value or force, and it is not usually part of what people take other retweeters to communicate. Like for acts of indicating, the communicative point of retweeting seems to be rather linked to their *ostensive* role: making the original content of the tweet available, directing one's followers' attention to it.

4 What retweets communicate, and how

4.1 Ostension and ostensive-inferential meaning

I have argued that retweeting is a form of quotation analogous to 'quoting-by-indicating', and that ostension plays a central role in determining the communicative value of retweets; but I have not yet explained *how* retweeters communicate by performing ostensive acts. Relevance theorists have argued that all intentional communication is ostensive – that every communicative act proceeds from an act of showing something to more complex communicative functions (Sperber and Wilson 1995, 49). The relevance-theoretic framework can be helpful to reconstruct the communicative mechanisms by means of which retweets acquire their meaning.

For relevance theorists, 'relevance' refers to the property of a stimulus that, when processed, brings about "the greatest contribution to the [receiver's] general cognitive goals at the smallest processing cost" (Sperber and Wilson 1995, 48). On this view, every act of ostension carries a tacit guarantee of relevance, which makes the communicative intention behind the ostension manifest, guiding interpretation. To produce a communicative stimulus (say, pointing at a billboard) is to make manifest an intention to make something manifest, thereby presenting it as worth processing (relevant). In turn, to understand what a communicative stimulus is intended to communicate is to *infer* what is it about the ostended content that the communicator assumed to be relevant (e.g. the fact that billboard contains information about what is on at the cinema, and that such information is useful to decide what to do tonight). This communicative process (understanding the meaning of a communicative act by reconstructing what would make it relevant) goes under the label of ostensive-inferential communication.

If retweeting involves some kind of digital ostension, communicating by retweeting should rely on inferential patterns like the ones just described. To see how this would work, let us consider **T6**.

(T6)

Donald J. Trump Retweeted



ProgressPolls @ProgressPolls · Aug 4

Who is a better President of the United States? #ObamaDay

61% President Trump

39% President Obama

In retweeting @ProgressPolls's graph, Donald Trump shows a poll to his followers: he performs an act of ostension by means of which he manifests an intention to make the poll manifest (that is, available to his followers) – an act comparable to indicating the poll. In doing so, he tacitly guarantees that the representation is relevant – that directing one's attention to the poll (and processing its content) is worth the effort. To understand what Trump is communicating through this act of ostension requires inferring why Trump assumed that this representation would indeed be relevant for his followers. A natural hypothesis is that he takes the information contained in the graph to be relevant because it is accurate, which may lead us to infer that he intends to communicate that he is a better President than Obama according to this recent poll. Alternatively, if the poll is still open, we may hypothesise that Trump rather takes this post to be relevant because people can still participate in it: in this case, we would infer that Trump is inviting his followers to take part in the survey, possibly to show their support by voting for him. These options are jointly available: Trump may be attempting to both claim that he is a better president according to the poll and invite people to take part in it. Note that on the proposed view, there need not be a univocal and precise answer as to what a given act of ostension communicates: ostensive-inferential communication is a fallible, open-ended process that need not have precise aims, and communicative success may be achieved to a higher or a lesser extent (Sperber and Wilson 1986, 59-60).

In sum, the interpretation of the communicative value of retweets is just like the interpretation of an act of indication: it does not rely on simple conventions (like the ones postulated by the endorsement hypothesis), but rather on complex and fallible inferential patterns, such as the ones that relevance theorists have attempted to model and describe.

Relevance and quotation

Like quotations, retweets involve a particular kind of ostension: ostension that is directed at a representation. As such, both communicative acts follow analogous patterns in how they can achieve relevance – patterns that have been explored in detail in the relevance-theoretic literature (e.g. Wilson 2012). Let us go back to the analogy with quotation, to explore some shared mechanisms through which quoting and retweeting can achieve relevance.

First, both quoting and retweeting can achieve relevance merely by fulfilling their attributive function. Although I have argued that this is rarely the point of issuing a retweet, relevance can nonetheless *in principle* be achieved in this way. That is, the point of showing a target tweet by retweeting it may simply be to

communicate that *someone* posted that tweet. For instance, commenting over T1 Donald Trump may insist that he merely intended to point out that *it was Lynn Thomas*, and not someone else, who claimed that the democrats are the only ones who are interfering in the US elections. This attributive component is what sometimes explains the linguistic data that we labelled as *deniability* and *negotiability*. Since retweets are always attributive, a user can always claim that she intended a retweet to achieve relevance *merely* in virtue of the attribution it communicates, and not by taking any stance towards its content.

Another way in which quotations can achieve relevance is by expressing the speaker's attitude towards the reported content – what relevance theorists call 'echoic uses' (Sperber and Wilson 1986, 237; Wilson 2012, §4.3) Echoic uses of quotation are a familiar phenomenon, whereby relevance is achieved by communicating the speaker's attitude towards the reported utterance. Consider the difference between (9), (12) and (15):

(9) Jenny said “The restaurant opens at 5 pm”

(12) As Jenny said, “The restaurant opens at 5 pm”

(15) Oh, and then Jenny said... “The restaurant opens at 5 pm”! [laughs]

Now, (9) merely reports what Jenny said. But (12) and (15) convey a specific attitude towards what Jenny said. The expression “as” in (12) signals that the speaker agrees with the reported content, whereas the laugh in (15) may signal (in the right context) that the speaker disagrees with the reported content. Similarly, a retweet can express agreement or disagreement (and further attitudes like surprise, discomfort, etc.) with the original tweet. Here relevance is achieved by communicating where the speaker stands towards the content of the retweet, either explicitly (by adding a comment that clarifies the retweeter's stance towards the content of the tweet) or implicitly (by means of other contextual cues). Amongst many possibilities, echoic uses include those retweets that communicate endorsement of the original tweet. In these cases, relevance is achieved by making agreement salient: the retweeter aims to communicate that her followers should pay attention to a tweet because she *agrees* with that tweet. While the endorsement view takes these cases to be the *only* communicative use of retweets, the present picture classifies them as *one* of the possible ways for retweets to achieve relevance.

4.3 Relevance and truth

Echoic uses explain how retweeters can express their stance towards the target tweet. But retweets are often meant to present the target tweet simply as *true*, rather than presenting it as something that the retweeter *takes to be true*. Something still needs to be said about the mechanisms through which speakers may present retweeted content as true, or even reassert it.

We face two challenges in this respect. First, as mentioned in §2.2, not all tweets are something that can be reasserted or presented as true, and it is not easy to draw a clear-cut, principled distinction between those that do and those that do not convey assertoric content. However, since my aim here is to reconstruct interpretative patterns (rather than conventions that apply to each and every case),

I will simply restrict my attention to retweets of tweets that an audience takes to convey some asserted content – call them *informative retweets*.

Second, unlike other approaches to the study of communication (paradigmatically, neo-Gricean theories), relevance theory places little emphasis on expectations of truthfulness, because it sees the exchange of truthful information as merely one of the many ways in which a communicative stimulus can achieve relevance in a context (Wilson and Sperber 2002; Van der Henst, Carles, and Sperber 2002). However, relevance theory does allow for expectations for truthfulness to play a central role in *some* communicative interactions. More specifically, whenever a stimulus is assumed to convey information about the world, that stimulus is relevant *only if* the content it conveys is true – and irrelevant if it is false (Sperber and Wilson 1995, 263-266; Wilson and Sperber 2002, 627, Sperber et al. 2010)¹⁵.

It is relatively easy to move from here to an account of retweets that does justice to the fact that informative retweets are sometimes interpreted as presenting their content as true. Informative retweets are (by definition) stimuli that convey content (the original tweet) that has been recognised as assertoric; in turn, assertoric content is propositional content that is presented as true, or that ‘purports’ to be true (for elaboration, Dummett 1973; Wright 1992, 23-34; Adler 2002, 274; Marsili 2018: 644-5). If a retweet is an ostensive act that makes a tweet manifest, carrying a tacit guarantee of its relevance, the result is that the retweeter is guaranteeing that some *content that purports to be true* is relevant. To make sense of this communicative act, we must ask: what is the most salient hypothesis about why the retweeter takes this content to be relevant? *Ceteris paribus*, the answer will be: the retweeter must take that content to do what it purports to do – namely, veridically represent a state of affairs. That is, lacking any contextual defeaters, the most salient hypothesis about why an informative retweet is taken to be relevant in a given context is that the retweeter takes its content to be true.

This should clarify why, from a relevance-theoretic perspective, informative retweets are often treated as reassertions, an aspect that was heavily stressed by the endorsement view. But there are important differences with the endorsement view, and that make the proposed explanation superior. A first difference (shared with the default-implicature version of the endorsement view) is that the relevance-theoretic account allows for contextual defeaters –like information about the retweeter’s personality, likely intention, and so forth– to cancel the assumption that any given retweet is meant to reassert or endorse the original tweet. This means that a retweet can at most *imply* that some content is true, avoiding the worries about cancellability, deniability and Moorean paradoxicality that affect the endorsement view (cf. §2.2). Second, the relevance-theoretic view does not postulate reassertion (or endorsement, for what matters) as the *primary* communicative function of retweets, but rather as one of the diverse ways in which retweets convey their

¹⁵ I am not attempting to imply that the relevance theoretic approach has the upper hand here. A neo-Gricean view can easily explain our expectations of truthfulness towards retweets by reference to the Cooperative Principle and the Quality Maxim (Grice 1989), in particular drawing on work that extends such expectations to implied content (A. Green 2017; Stokke 2018). My aim here is merely to outline how the relevance-theoretic framework adopted so far can explain the epistemic expectations engendered by informative retweets.

meaning. Unlike the endorsement view, the relevance-theoretic proposal can therefore easily explain the communicative value of non-sentential retweets (retweets of images, videos, music, etc.), where nothing is asserted, but some content is nonetheless presented as worthy of the audience's attention.

This concludes the analysis of the communicative value of retweets. All retweets are acts of ostension that make a tweet available to an audience, presenting it as worthy of cognitive engagement. They do not have a single communicative function, such as endorsing the target tweet. Their communicative import is much richer, and derivable inferentially, so that a relevance-theoretic model can be fruitfully applied to offer a systematic explanation of the many communicative goals that retweets can achieve over and beyond making the original tweet manifest (the ostensive act). The variety of ways in which retweets achieve their communicative goals is summarised in Figure 1.

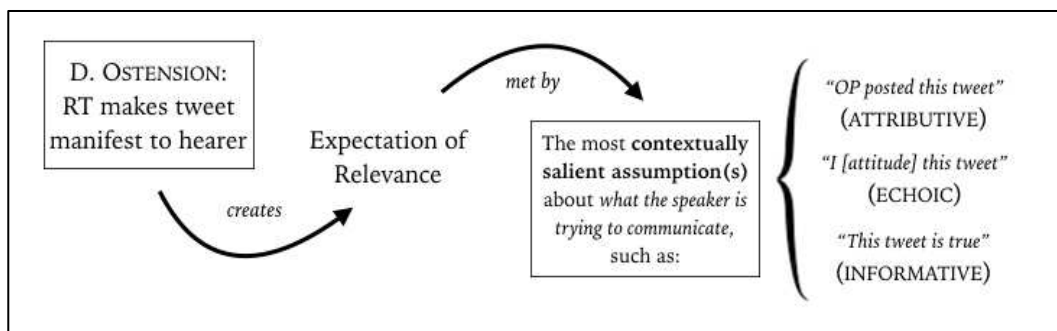


Figure 1: A representation of how retweets achieve their communicative value

4 The epistemic value of retweets

5.1 Epistemic expectations: on truthfulness and accountability

What about our expectations of sincerity and accuracy about retweets – that is, *prima facie* expectations that people repost informative content only if it is accurate and believed to be accurate? To some extent, they parallel the expectations we have towards reported content in quotation. We have seen (§2.3) that reporting what someone said can be taken to imply that what they said is true. For instance, if you ask me when the trattoria opens and I reply with (10), I would be taken to imply that it likely opens at 5pm, although I did not assert it. If it turns out that I know that the restaurant opens much later (or that I was aware that Jenny was almost surely lying) I cannot strictly speaking be criticised for having lied (Stokke 2013), but I can surely be criticised for misleading – that is, for purposely conveying a proposition that I believe to be false, in order to deceive my audience (cf. Lackey 2007, 614-8; Saul 2012; A. Green 2017).

Similar criticisms are available against people who retweet content that they believe to be false (or have no good reason to believe to be true), with the intention of passing it as reliable information. Retweeters of false or deceptive content are

liable to be criticised for presenting as relevant material that is not (as relevance requires veracity, at least for content that purports to be informative), and for having purposefully engaged in deceptive communicative behaviour.

Like a misleader who deceives without lying, the deceptive retweeter enjoys a certain level of deniability. They can always insist that they did not mean to imply anything about the veracity of what they reposted, or claim that they merely posted it to invite discussion, or to point out that the OP had said something false. Still, whenever these excuses are insincere, the retweeter's communicative behaviour is liable to be criticised for purposefully attempting to deceive someone. In other words, the moral and epistemic liability of a deceptive retweeter is equivalent to that of a misleader who attempts to deceive without lying¹⁶.

Against this diagnosis, it may be observed that users indeed face *severe* social sanctions for what they retweet, so that my proposal fails to capture in full the epistemic responsibilities undertaken by retweeters. Just consider the effects of retweets issued by public figures, who can be harshly criticised for what they retweet. To mention a few cases from US politics, in 2018 Donald Trump retweeted an Islamophobic video from the neofascist association Britain First, Trump's associate Dinesh d'Souza a message containing a #burnthejews hashtag, and Rudy Giuliani a message describing George Soros as the "anti-Christ". They all have been under fire for these retweets, each has been pressured into apologizing publicly, and their political reputation has suffered because of what they reposted. These cases are not exceptional: there are countless documented cases of individuals being asked to resign from their positions, fired and even sued for what they retweeted¹⁷. We can conclude that people are socially responsible (and can face significant social sanctions) for what they retweet.

These observations, however, do not contradict the picture that I have just outlined. Clearly, one can face severe social sanctions for *implying* that something false is the case: fraud (deception that intentionally causes harm) violates civil law, and when detected may well determine that the perpetrator be criticised, pressured to apologise, fired or even sued. Furthermore, speech can violate social norms without being assertoric or violating any epistemic expectation – so that one cannot infer assertoric force (and, consequently, lie-aptness) from the social sanctionability of a speech act. For instance, using a slur to report someone else's speech can justly be interpreted as an instance of hate speech itself, eliciting the relevant criticism and social sanctions (Schlenker 2007; Cepollaro, Sulpizio, and Bianchi 2019), even

¹⁶ For an account of the moral and epistemic vices of which the misleader is culpable, and a comparison with lying, see e.g. Adler (1997), Saul (2012), Pepp (2018), Marsili (2019). Note, further, that (both in retweeting and misleading) what constitutes a *plausible* denial may vary from context to context. We know from experience that it is easier to backpedal on mundane matters than it is about matters of life and death, that claims made by public figures and high officials are held against higher standards of accountability than those made in informal conversations between friends, and so forth (for elaboration, see Pinker 2007; Lee and Pinker 2010; Camp 2018).

¹⁷ Two lawsuits concerned libellous retweets gained the attention of the media, both in 2012: one concerning a [Swiss journalist](#), and another involving Alistair MacAlpine, whose legal case that could have led to the [largest number of defendants](#) (about 9000) in British history. An example of an employee being fired over a retweet is discussed in Rosemary Bennett, 'Student editor Angelos Sofocleous fired in transphobia row', *The Times*, 21/9/2018. See also Stephanie Gruner Buckley, 'Want to face a libel suit? Just hit "retweet"', *Quartz*, 20/11/12.

when it is conveyed by an utterance that reports the offensive content without asserting it, such as:

(16) John told me that you are a *****

To recapitulate. I have argued that informative retweeting carries a *defeasible* presumption that its content (the original tweet) is relevant because it is truthful and accurate. It follows that uncommented retweets at most *imply* that something is the case, and are therefore liable to the criticisms one would face for *misleading*, as opposed to lying. This should not be confused with the claim that retweeters are in *no way* accountable for what they retweet, nor with the claim that they should not be held against them. On the contrary, retweets can be deceptive and therefore epistemically blameworthy; but they are criticisable because they violate the epistemic norms governing what is suggested or insinuated, rather than those regulating what is asserted.

5.2. Cheap rumours: on the dangers of retweets

I have mentioned that retweets play a fundamental role in the spread of misinformation online. In contributing to give visibility to a piece of information, retweets (and shares on social media more generally) are one of the main vectors for the diffusion of false information on the internet (Vosoughi, Roy, and Aral 2018). The way misinformation spreads via retweeting cascades has been compared in the scientific literature to the way rumours propagate in offline communication (Friggeri et al. 2014). In both rumour-mongering and retweeting, people contribute to the diffusion of an unverified claim by reporting that that claim has been made (that is, by quoting someone). Both are by definition second-hand reports: a piece of testimony is a rumour only if it involves reporting a claim someone else has made, rather than first-hand knowledge (Coady 2012). But rumours and informative retweets also differ in a number of ways, and outlining their dissimilarities can be helpful to highlight the particular challenge that retweets present to our current epistemic environment.

An important difference between retweets and rumours is their *cost*. Reporting what someone has said (a rumour) in ordinary language has cognitive and practical costs. One needs to remember or otherwise mentally access the target utterance (the rumour), then determine which features of that utterance are salient and should be included in the reported utterance, and finally physically produce the utterance that reports the original claim (by vocalising it, writing it, or otherwise producing it).

These practical and cognitive costs are surely low, but still comparatively high when one considers the simplicity of retweeting information. Uncommented retweeting merely requires clicking on the ‘retweet’ icon (plus a second click on ‘retweet’ in the resulting popup window). It does not require recollecting and reformulating the original claim, and is extremely fast and simple to perform. Retweeting is thus an extremely economic and rapid way of communicating, and this is arguably the reason for its growing prevalence on social media. According to a study by Liu, Kliman-Silver, and Mislove (2014), retweets accounted for over 25% of the material posted on Twitter in 2014, and the number had been steadily

growing since 2009, when native retweets were introduced¹⁸. Communication on social media thus differs from offline communication in that in the former environment it is far more common to communicate something by reporting what someone else has communicated. While about one in four utterances is a form of reported speech on Twitter, nothing comparable is true of offline communication. Understanding the significance of the cost and prevalence of retweets allows us to fully appreciate the epistemic challenge they represent to the flow of reliable information on online platforms. Retweets (or at least *informative* retweets) are insidious because they are subjects to milder expectations than the ones attached to explicit claims. Retweeters enjoy more deniability: they can always insist that they did not mean to imply that the original tweet was reliable or veridic. These lower epistemic standards (in turn linked to lower reputational costs for retweeting false information), together with the lower practical costs¹⁹ linked to producing a retweet, contribute to retweets being often produced without paying too much attention to the accuracy of the information that is thereby spread: a research by Gabielkov et al. (2016) estimates that 15% of the links shared on Twitter gets retweeted without being read.²⁰

Humans rely heavily on information that circulates by being communicated. As our communicative environment changes, so change the challenges that we face as we strive to access reliable information, information that is fundamental to make our decisions and to form an accurate picture of the world around us. The introduction of retweets on Twitter (and of analogous features on other social media platform) has had a dramatic impact on the communicative environment of social media platforms. This paper has offered a systematic account of how they work, linking their communicative import to a peculiar species of quotation, *digital ostension*, that can achieve different communicative functions. It has argued that it is primarily because of their low cost, their prevalence and the limited epistemic responsibilities to which they are subject, that they represent a particularly pernicious challenge to contemporary society and deliberative democracies, whose health heavily depends on flow of truthful and reliable information.

¹⁸ At the time, written retweets (using the syntax “RT @OP: *p*” described in footnote 9) accounted for less than 5% of all retweets, indicating that their growing prevalence is indeed likely to have been determined by a sudden decrease in their practical and cognitive cost.

¹⁹ Arguably, also these non-reputational costs affect the communicator’s choice to carefully sticking to truthful signals, as suggested by the role that physical ‘handicaps’ seem to play in maintaining animal signalling systems reliable throughout generations (Zahavi and Zahavi 1997; Searcy and Nowicki 2005)

²⁰ Some clarification is needed. The study (Gabielkov et al. 2016, §2.3) found that 59% of the links (to news sites) shared on Twitter are never clicked *after* they have been shared. The media have repeatedly misreported this result; for instance, a viral article by Caitlin Dewey (*Washington Post*) titles “6 in 10 of you will share this link without reading it, a new, depressing study says”. Both the headline and the article are inaccurate, since the figure identified by Gabielkov and colleagues merely indicates that 59% of the links posted on Twitter remain ‘silent’, or unclicked. That being noted, since several of these unclicked links get retweeted, we can deduce that a subset of that 59% has been retweeted without having been read. Personal communication with the authors revealed that they estimate that subset to be 15% of the total traffic: in other words, it is only 15% of retweets (of *news links*, excluding retweets by bots) that aren’t read before being shared.

Acknowledgments

I would like to thank Deborah Mühlebach, Jennifer Saul, Josep Macià, Don Fallis, Josie McQuillan, Marina Sbisà, Maciej Witek and Patrick Connolly for helpful comments on earlier versions of this draft. This paper has been presented at the *Perspective on Speech as Action* workshop at the University of Trieste, the *National Meeting of the Italian Society for Philosophy of Language* at the University of Cagliari, the *Philosophy of language seminar* at the University of Sheffield, and the *JeSyP 2019* workshop at the University of Jena. I am grateful the audiences of all these events for helpful comments and feedback.

References

- Adler, Jonathan E. 1997. Lying, Deceiving, or Falsely Implicating. *Journal of Philosophy* 94 (9): 435–52.
- . 2002. *Belief's Own Ethics*.
- Allcott, Hunt, and Matthew Gentzkow. 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31 (2): 211–36.
- Alston, William P. 1964. Linguistic Acts. *American Philosophical Quarterly* 1 (2): 138–46.
- . 2002. Illocutionary Acts and Linguistic Meaning. In *Foundations of Speech Act Theory*, 37–57. Routledge.
- Austin, John Langshaw. 1962. *How To Do Things With Words*. Oxford: Clarendon Press.
- Boyd, Danah, Scott Golder, and Gilad Lotan. 2010. Tweet , Tweet , Retweet : Conversational Aspects of Retweeting on Twitter. In *Proceedings of the 43rd Hawaii International Conference on System Sciences*, 1–10.
- Camp, Elisabeth. 2018. Insinuation, Common Ground, and the Conversational Record. In *New Work on Speech Acts*, edited by Daniel Fogal, Daniel W. Harris, and Matt Moss. Oxford University Press.
- Caponetto, Laura. 2018. Undoing Things with Words. *Synthese*, no. November 2017. Springer Netherlands.
- Cappelen, Herman, and Ernie Lepore. 1997. Varieties of Quotation. *Mind* 106 (423): 429–50.
- Carr, Caleb T., David B. Schrock, and Patricia Dauterman. 2012. Speech Acts Within Facebook Status Messages. *Journal of Language and Social Psychology* 31 (2): 176–96.
- Cepollaro, Bianca, Simone Sulpizio, and Claudia Bianchi. 2019. How Bad Is It to Report a Slur? An Empirical Investigation. *Journal of Pragmatics* 146. Elsevier Ltd: 32–42.
- Clark, Herbert H, and Richard J Gerrig. 1990. Quotations as Demonstrations. *Language* 66 (4): 764–805.
- Coady, David. 2012. Rumour Has It. *International Journal of Applied Philosophy* 20 (1): 41–53.
- Davidson, Donald. 1964. Quotation. In *Words and Objections*, edited by Donald Davidson and Jaakko Hintikka, 21:307–8.
- Dayter, Daria. 2014. Self-Praise in Microblogging. *Journal of Pragmatics* 61. Elsevier B.V.: 91–102.
- Dummett, Michael. 1973. *Frege: Philosophy of Language*. Vol. 49. Duckworth.

- Fleischmann, Suzanne. 1999. Pragmatic Markers in Comparative Perspective: A Contribution to Cross-Language Pragmatics. In *PRAGMA 99 Proceedings*.
- Friggeri, Adrien, La Adamic, Dean Eckles, and Justin Cheng. 2014. Rumor Cascades. *Icwsn*, 101–10.
- Gabielkov, Maksym, Arthi Ramachandran, Augustin Chaintreau, and Arnaud Legout. 2016. Social Clicks: What and Who Gets Read on Twitter? *ACM SIGMETRICS Performance Evaluation Review* 44 (1): 179–92.
- Goodman, Nelson. 1974. On Some Questions Concerning Quotation. *The Monist* 58 (2): 294–306.
- Green, Adam. 2017. An Epistemic Norm for Implicature. *Journal of Philosophy*.
- Green, Mitchell S. 2017. Assertion. *Oxford Handbook Online*.
- Grice, H. P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Henst, Jean-Baptiste Van der, Laure Carles, and Dan Sperber. 2002. Truthfulness and Relevance in Telling The Time. *Mind and Language* 17 (5): 457–66.
- Labinaz, Paolo, and Marina Sbisà. 2020. Speech Acts and the Dissemination of Knowledge in Social Networks. In *Internet Pragmatics: Theory and Practice*, edited by Haberland, Yu, and Xie. Benjamins Publishing.
- Lackey, Jennifer. 2007. Norms of Assertion. *Noûs* 41 (4): 594–626.
- Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, et al. 2018. The Science of Fake News. *Science* 359 (6380): 1094–96.
- Lee, James J., and Steven Pinker. 2010. Rationales for Indirect Speech: The Theory of the Strategic Speaker. *Psychological Review* 117 (3): 785–807.
- Liere, Diederik Van. 2010. How Far Does a Tweet Travel? Information Brokers in the Twitterverse. *Proceedings of the International Workshop on Modeling Social Media, MSM '10*, 1–4.
- Liu, Yabing, C Kliman-Silver, and Alan Mislove. 2014. The Tweets They Are A-Changin': Evolution of Twitter Users and Behavior. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 305–14.
- Marsili, Neri. 2018. Truth and Assertion: Rules versus Aims. *Analysis* 78 (4): 638–48.
- . 2019. Immoral Lies and Partial Beliefs. *Inquiry* 0 (0). Taylor & Francis: 1–11.
- Metaxas, Panagiotis Takis, E. Mustafaraj, K. Wong, L. Zeng, M. O'Keefe, and S. Finn. 2015. What Do Retweets Indicate? Results from User Survey and Meta-Review of Research. *Proceedings of the 9th International Conference on Web and Social Media, ICWSM 2015*, 658–61.
- Metaxas, Panagiotis Takis, and Eni Mustafaraj. 2013. The Rise and the Fall of a Citizen Reporter. *Proceedings of the 5th Annual ACM Web Science Conference, WebSci'13*, 248–57.
- Metaxas, Panagiotis Takis, and Twittertrails Research Team TTRT. 2017. Retweets Indicate Agreement, Endorsement, Trust: A Meta-Analysis of Published Twitter Research. *ArXiv Preprint*.
- Nic, Newman, David A. L. Levy, and Rasmus Kleis Nielsen. 2018. Reuters Institute Digital News Report. *University of Oxford*.
- Page, Ruth. 2014. Saying “Sorry”: Corporate Apologies Posted on Twitter. *Journal of Pragmatics* 62. Elsevier B.V.: 30–45.

- Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin UK.
- Pepp, Jessica. 2018. Assertion, Lying, and Untruthfully Implicating. In *The Oxford Handbook of Assertion*, edited by Sanford C. Goldberg. Oxford University Press.
- Pinker, Steven. 2007. The Evolutionary Social Psychology of Off-Record Indirect Speech Acts. *Intercultural Pragmatics* 4 (4): 437–61.
- Recanati, Francois. 2000. Oratio Obliqua, Oratio Recta. An Essay on Metarepresentation.
- Rini, Regina. 2017. Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal* 27 (2S): E-43-E-64.
- Rossholm, Göran. 2004. *To Be and Not to Be: On Interpretation, Iconicity and Fiction*. Peter Lang.
- Saka, Paul. 1998. Quotation and the Use-Mention Distinction. In *Mind*, 107:367–87.
- Saul, Jennifer. 2012. *Lying, Misleading, and The Role of What Is Said*. Oxford University Press.
- Schlenker, Philippe. 2007. Expressive Presuppositions. *Theoretical Linguistics* 2007 (0617316): 1–6.
- Searcy, William, and Stephen Nowicki. 2005. *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems*. Princeton University Press.
- Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Vol. 626. Cambridge university press.
- Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The Spread of Low-Credibility Content by Social Bots. *Nature Communications* 9 (1). Springer US: 4787.
- Shearer, Elisa, and B Y Jeffrey Gottfried. 2017. News Use across Social Media Platforms 2017. *Pew Research Center*, 17.
- Silverman, Craig. 2016. This Analysis Shows How Fake Election News Stories Outperformed Real News on Facebook. *BuzzFeed News*.
- Sperber, Dan, and Deirdre Wilson. 1995. *Relevance Theory*. 2nd ed. Blackwell Publishing.
- Stokke, Andreas. 2013. Lying, Deceiving, and Misleading. *Philosophy Compass* 8 (4): 348–59.
- . 2018. Lying, Sincerity, and Quality. In *The Oxford Handbook of Lying*, 26–29.
- Sunstein, Cass R. 2001. *Echo Chambers: Bush v. Gore, Impeachment, and Beyond*. Princeton University Press Princeton, NJ.
- Vosoughi, Soroush, and Deb Roy. 2016. Tweet Acts: A Speech Act Classifier for Twitter. *ICSWM Proceedings*, no. ICWSM: 711–14.
- Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. The Spread of True and False News Online. *Science* 359 (6380): 1146–51.
- Walton, Kendall L. 1984. Transparent Pictures: On the Nature of Photographic Realism. *Critical Inquiry*, no. 11.
- Wikström, Peter. 2014. Srynotfunny: Communicative Functions of Hashtags on Twitter. *SKY Journal of Linguistics* 27: 127–52.

- Williams, Bernard Arthur Owen. 2002. *Truth and Truthfulness An Essay in Genealogy*.
- Wilson, Deirdre. 2012. Metarepresentation in Linguistic Communication. In *Meaning and Relevance*, 230–58.
- Wilson, Deirdre, and Dan Sperber. 2002. Truthfulness and Relevance. *Mind* 25: 1–41.
- Wright, Crispin. 1992. *Truth and Objectivity*. Cambridge, MA: Harvard University Press.
- Zahavi, A., and A. Zahavi. 1997. *The Handicap Principle: A Missing Piece of Darwin's Puzzle*. Oxford: Oxford University Press.