The Philosophy of Online Speech

For the vast majority of human history, linguistic communication has been carried out through face-to-face conversations in which two or more speakers share an environment, have access to each other's non-verbal cues, and use the traditional (i.e., visual or auditory) modalities of speech. In recent years, online speech has become not only a supplement to verbal communication but, for some, the dominant means by which they communicate with some research indicating that, for Gen Z and Millenials, a majority of their communication occurs digitally rather than face-to-face. This kind of revolution in the primary vehicle of language use is unprecedented in human history and its social repercussions are far from understood. This chapter critically examines the idea that online communication gives rise to the emergence of unique and novel speech acts. First, it examines the claim that online speech occurs in a unique kind of communicative context that is sufficiently distinct from offline communication to require different theoretical tools to be examined. Second, it assesses the claim that online communication provides for the development of novel speech acts that are not found in offline communication, with a particular focus on sharing and liking. While examining recent philosophical work in the field of online speech, the chapter argues that many of the apparently novel features of online speech arise through attempts to recreate features of offline communication, in particular, non-verbal cues, in a digital context. Nevertheless, several distinctive features of online communication are identified. Finally, it is argued that what is most distinctive about online communication is not how it alters the range of speech acts we perform but why we choose to engage at all.

FOR the vast majority of human history linguistic communication has been carried out through face-to-face conversations in which two or more speakers share an environment, have access to each other's non-verbal cues, and use the traditional (i.e., visual or auditory) modalities of speech. In recent years, technologically mediated speech has become not only a supplement to verbal communication but, for some, the dominant means by which they communicate. This kind of revolution in the primary vehicle of linguistic behaviour is unprecedented and its social repercussions are far from understood. People now spend hours each day communicating across screens (Wagner et al. 2021), an increasing portion of education now takes place in front of a computer, while large number of couples are meeting and falling in love online or through apps (Rosenfield 2019). This has coincided with the decline of physical 'third places' more generally and the ongoing effects of the COVID-19 pandemic. Social media, the locus of much of this communication, is deeply woven into the fabric of modern life. A 2023 Gallup survey suggests that the average American teenager spends 4.8 hours on social media every day although much of this time is spent on YouTube and TikTok rather than text-oriented platforms (Rothwell 2023). The aim of this chapter is to determine how revolutionary these developments really are for our linguistic behaviour. Specifically, it will provide a broad survey of recent work on the philosophy of online speech with an eye on two questions that are often left implicit: To what extent are the developments enabled by online communication genuinely *novel*? And to what extent can the various features of any new linguistic acts be explained by appeal to the particulars of the technological infrastructure of online media platforms?

The question of novelty admits a spectrum of possible answers. A revolutionary view of online speech sees it as a radical upheaval of linguistic practice. According to this view, the internet hasn't just changed the scale of communication, connecting countries and audiences but, in the words of Lucy McDonald, "the internet

and especially social media, has also changed *what it means* to communicate" (McDonald 2021 - emphasis in original). The internet provides a vehicle for new speech acts (e.g., 'liking', retweeting, subtweeting), new kinds of speaker (e.g., chatbots, sock-puppets, trolls), and new vehicles of meaning (e.g., emoji, gifs, reactions) that are *sui generis*. One consequence may be that philosophers and linguists will require new theoretical tools to study online speech. In contrast, a more conservative position holds that the internet merely amplifies certain familiar aspects of communication but much of what appears novel, for example, emoji or likes, are merely attempts to reconstitute familiar features of face-to-face communication such as gesture and body language within a new context. The internet may have allowed for communication on a greater scale, but it doesn't introduce anything linguistically new.

The second question concerns the relationship between linguistic behaviour and the particular infrastructures that facilitate it. Research in this area has taken place under the shadow of technological determinism.[1] While it is rarely claimed that our communications technology forcibly determines the significance of our linguistic practices, it is sometimes argued that it provides particular 'affordances' for linguistic behavior (Tufekci 2017; Davis 2020; Settle 2021; Tanesini, 2022, 2025). While the term 'affordances' is contested, here they'll be understood as the range of actions the technology facilitates or makes possible. By making some actions easier than others to perform and introducing further incentives for certain actions, a communication platform may influence users' behaviour.[2] This chapter will discuss these ideas in terms of *media ideology* (Gershion 2010; Record and Miller 2022). A media ideology is a system of beliefs about how a medium structures and communicates a message (Gershon 2010). These beliefs shape how users interpret the messages conveyed on a platform. While there may not be consensus on the media ideology of a given platform, ideologies must be at least partially shared if communication is to succeed. The beliefs comprising a media ideology are not confined to context and content but also include wider pragmatic and normative considerations. For example, an email is typically understood to be more formal than a text while a DM (direct message) will be interpreted differently from a publicly posted message. Some users may approach a platform to disseminate research and news while others use it for 'shit-posting' or disinformation. Much contemporary work on online speech is the analysis of media ideologies; are retweets endorsements? What is the significance of 'liking' a post? What does 🫠 mean? As new norms and patterns of linguistic behaviour emerge alongside changes to the technological infrastructures supporting them, it is important to examine the extent to which these developments are driven by the technological changes. How, if at all, is media ideology determined by the infrastructure of internet platforms? The answers to these questions have a direct impact on how we should respond to the social problems that are often blamed on the nature of online communication. For example, it has been argued that social media platforms contribute to political radicalisation and encourage people to engage in epistemically harmful behaviour (Rini 2017; Nguyen 2020,

---

[1] Consider Ong on the shift from 'oral' to 'written' culture (Ong and Hartley 2013), Eisenstein on the shift from 'manuscript' to 'print culture' (Eisenstein 1979), or McLuhan's claim that 'the medium is the message' because it is the medium that shapes and controls the scale and form of human association and action' (McLuhan 2017).

[2] Succinctly stated "The effects of using any particular social media platform result from the combination of the features of the site, called affordances, and the norms that develop around the way in which users employ those affordances… certain technological features foster particular psychological phenomena" (Settle 2018: 8).

2021; O'Neill 2022; Benton and Schmidt 2024; Tanesisi, 2025). If these social problems emerge as a by-product of our communication infrastructure itself, then a structural intervention is required in order to address them. Alternatively, if our norms of communication are relatively independent of their vehicle, then we may be able to rely on interventions that leave the platform infrastructure intact, perhaps focussing on making people more virtuous users of these tools (Frost-Arnold 2021; Bail 2021). These positions are obviously not exclusive but the details of our interventions should be informed by an analysis of the cause of the problem.

In what follows, I'll use the term 'platform' for specific social media websites (e.g., Bluesky, 'X' formerly Twitter, Facebook, Reddit, Instagram, etc.) and 'infrastructure' for how they are designed — the expressive options provided to users, how posts are presented, the algorithms governing what is presented, etc.,.[3] I'll typically use the term 'posting' for contributing text to a platform and 'sharing' for replicating that post in a new context. The focus in what follows will largely be on online conversations, that is, conversations that take place on social media platforms rather than private messaging services although discussions that don't make explicit appeal to the public nature of online speech should carry over.

The structure of his paper will work outward from the notion of context, to the speech acts online contexts enable, and finally to the incentives for behaviour these create. Section 1 will focus on context both as something restricted by pragmatic impoverishment and as something enlarged by the increased potential for audiences in online communication. Section 2 will give an overview of work on distinctively online speech acts, namely, posting, reposting, liking. Section 3 looks at the claim that platform infrastructure plays a role in shaping discursive goals. As the philosophical investigation of online speech is a field in its infancy, this chapter will also try to provide as thorough an overview of current work as possible as well as a sketch of this work relates to the general theme of media ideology.

28.1. Context

All linguistic activity occurs within a determinate context. Research on the structure of online contexts has primarily been driven by epistemic rather than semantic questions, but it has resulted in some consensus about the unique properties of online contexts (Bianchi 2014). One of the first questions facing philosophers of online speech is how this context should be analysed as the philosophy of language furnishes several alternative models of context. Lewis treats context as, "a location — time, place, and possible world - where a sentence is said" (Lewis 1979) and while this notion of context is sufficient to fix the reference of indexicals, it doesn't provide a deeper specification of the pragmatic cues available to interpreters and so it is

[3] There have been several attempts to identify a set of core features that distinguish online from offline speech. Record and Miller identify six stable features of posts: modularity — posts exist as discrete units, algorithmic surfacing – algorithms often occluded from the user determine which posts are visible to users, adjacency — posts are placed adjacent to each other regardless of content, archivability, — posts may surface long after they were originally made and independent of their original context, modifiability — users can interact with posts, and accessibility (Record and Miller 2022). Others have highlighted that posts are searchable and permanent although these features are not universal. While many users post under their own names with the platform infrastructure preserving these posts, on sites like 4chan, users are almost exclusively anonymous while the platform is ephemeral with old threads being deleted (Bernstein et al. 2011).

of limited use for examining the unique features of online communication. The alternative, more influential, conception of context appeals to what the conversational participants take to be common ground; what they accept for the purposes of conversation and believe that others accept… and so on. This version of common ground emphasizes what is epistemically accessible to conversational participants during an exchange. In most cases, this includes the identities of those who are party to a conversation and any goings on in the environment ('manifest events') such as the weather or salient objects and occurrences (Stalnaker 1978 [1999]). We can assume that in offline communication, the common ground also includes mutual beliefs about non-verbal cues such as eye gaze, nods, smiles, hand gestures, whereas for online communication it includes any features of media ideology that inform the interpretation of speech acts performed on the platform. In the ideal case, users mutually believe that 🤮 has a mutually agreed meaning. In this section, I'll highlight some of the ways in which online contexts are pragmatically impoverished, in particular, the loss of non-verbal cues and the effects of 'context collapse'. We will look first at some pragmatic features of online contexts before discussing some semantic considerations.

28.1.1 Pragmatic Impoverishment: Non-Verbal Cues

Offline communication is typically accompanied by a rich array of non-verbal cues including facial expressions, eye contact, posture, and gesture, as well as information conveyed by prosody and the loudness and speed of speech. None of this occurs by default in online, text-based communication. Focussing first on prosody, while it is possible to indicate emphasis *to some extent* textually, text is not typically used with the full expressive resources of prosody. In offline communication, prosody, i.e., the pattern of stress and intonation in an utterance, is an important marker of information structure (Beaver et al. 2017). To take a simple example, the sentence 'I didn't pet the cat' can convey several different propositions depending on which expressions are prosodically focussed.

(i) *I* didn't pet the cat

(ii) I *didn't* pet the cat

(iii) I didn't *pet* the cat

(iv) I didn't pet *the cat*

While this example illustrates that focus may still be conveyed by font (a point highlighted in Keisler, 2021), such ambiguities are typically resolved in offline communication by prosodic focus. To apply such focus in text, whether through **bolding**, font change, or *italics*, requires conscious effort on the part of the speaker while these options may not be available on certain platforms. The loss of this information can have several consequences. For concision, I will confine myself to three.

The first is that information structure is often taken to indicate what the speaker takes to be the question-under-discussion in an exchange (Roberts 2012). The utterance (i) is best understood as a response to the

question ‘who pet the cat?’, while (iii) is a response to ‘what was done to the cat?’. According to the Stalnaker-Roberts model of conversation dynamics, the question-under-discussion determines the felicity of further assertions and assertions are expected to be relevant to a particular question. The loss of this structure in online exchanges may play a role in generating fruitless disputes about the relevance of certain claims to presupposed questions-under-discussion as well as failures to identify implicatures that rely upon violations of relevance. Furthermore, the asynchronous nature of online communication, including the fact that posts can be copied and pasted in new contexts, ensures that it may not be possible at all to determine the question or subject matter to which the speaker took their contribution to be a response. While some platforms offer users an official QUD (Facebook has traditionally asked, ‘what’s on your mind?’ while Twitter asked ‘what’s happening?’), these don’t appear to play a role in structuring the discourse that follows meaning that, as Sanford Goldberg puts it, “posts can have the feel of public announcements rather than contributions to a conversation itself” (Goldberg 2021 — see also Tirrell on ‘hit-and-run speech acts’, Tirrell 2025).

The second consequence is potentially more serious. If a recent trend in the philosophy of language is correct, the QUD plays a role in determining the semantic content of an utterance in the case of knowledge ascriptions (Schaffer and Szabó 2014), exculpature (Hoek 2018), or simply determining ‘what is said’ (Schoubye and Stokke 2016). If these theories are correct, the difficulty of determining the QUD in an online exchange can make it harder for interpreters to work out the content of the claims people make online (Pollock 2024).

Third, the loss of prosodic information can undermine users’ ability to engage in certain kinds of speech. McMullen details how the loss of paralinguistic cues undermines our capacity to be sarcastic online (McMullen 2025). Offline, sarcasm can be identified by inverse pitch obtrusion, ‘flattened affect’, and a saccharine singsong (Haiman 1998). These features are absent from online contexts, preventing a speaker from signalling to interpreters that what they are saying should be understood insincerely.[4] This can contribute to the misunderstandings often discussed in reference to online communication. While certain emojis, e.g., 🙃, can provide a cue that one is being sarcastic, the cost may be the humour.

28.1.2 Context Collapse

Non-verbal cues also serve as a means of indicating who is taking part in a conversation. In offline contexts, participants can tell who is listening or intending to contribute to a discussion by tracking body language and facial expressions. In contrast, online speakers often have no means to track who is aware of their contributions. Sanford Goldberg argues that, without a means to track the participants in a conversation, speakers can’t track whose beliefs are generating the common ground (Goldberg 2021). This suggests that online conversations take place in a structurally defective context, that is, the infrastructure which makes online conversations possible ensures that participants will be unable to assume that others are presupposing what they presuppose. Meanwhile, Onora O’Neill has argued that anonymity, a feature of some platform

[4] Rini argues that pop-up communities can form around prominent users that allow for their own specific norms and in-jokes which can help with this (Rini 2025). Few philosophers would have trouble understanding classic works by @lastpositivist.

infrastructures, undermines the capacity to determine which norms of communication are being met (O'Neill 2022). Epistemological work on online communication largely focusses on defective epistemic networks (as manifest in work on epistemic bubbles, echo chambers, conspiracy theories, etc.). A central focus of this work has been the phenomenon of 'context collapse'. Context collapse is the "blurring or merging of multiple contexts or audiences into one" (Frost-Arnold 2021). This can either be intentional and with the consent of the speaker, "context collusion", or without the speaker's consent, as when a message is screenshot and shared, "context collision" (Davis and Jurgenson 2014). Facebook, for example, connects one to friends, co-workers, family members, and other distinct groups with different interpretive resources among whom a speaker may maintain a range of different identities.

Linguistically, this becomes relevant when an utterance or post that is produced in one context is picked up in a context where different interpretive resources are available. Frost-Arnold argues that online context collapse promotes misunderstanding by enabling utterances to be shared outside of their original context. As a result, it can be difficult to recover the proposition expressed (Frost-Arnold 2021). Joey Pollock has argued that this makes it difficult to burst an epistemic bubble merely by presenting evidence to a person as one will almost always lack the interpretive resources required to interpret speakers from outside one's own, local epistemic community (Pollock 2024). Lucy McDonald proposes that we understand context collapse, not as a merging of different contexts in the sense of audience aggregation, but as an overlapping of different Lewisian language games, each with its own distinct scoreboards and evolving norms of communication (McDonald 2025). According to this view, to engage in online conversations is to simultaneously be bound by overlapping, different, and possibly conflicting communicative norms. McDonald argues that this is different in scale and kind from traditional media context collapse as all participants are both broadcast producers and consumers.

The opportunity to speak to multiple audiences at once makes it easier to engage in certain kinds of deceptive speech. Work on dogwhistles (Saul 2018) and insinuation (Camp 2018) sheds light on how speakers negotiate communication on a platform with multiple contextually determined audiences. The public nature of speech on social media platforms makes it easier for anyone to speak across contexts. For example, Woodard and Neufeld argue that platforms like Twitter, by facilitating the easy emergence of alternative audiences, in- and out-groups, and by having unclear conditions on minimal relevance (described above), are well-placed to facilitate practices of insinuation (Woodard and Neufeld 2025). One reason for this is that if one is 'called out' for a claim one appears to make, it is easy to point to an alternative audience for whom the literal rather than the insinuated content of an assertion is more natural (a pattern also seen in meme usage, Ernst 2025). This doesn't mean that subtweeting is a uniquely online phenomenon - one could, in theory, write an academic paper that possesses all the properties of a subtweet, but it does indicate an important way in which the context of online discourse makes it easier to perform certain kinds of speech act.

28.1.3 Indexicals

Finally, we can examine the media ideologies of online platforms and the understandings of context they generate by considering the interpretation of indexical expressions online. In November 2021, the K-pop star j-hope posted 'Love U💜' on Twitter (receiving 2.9 million likes as a result). As this was an independent post rather than a reply, the 'U', context does not furnish us with a unique referent. There are several ways we can think about how the referent is determined. It has been suggested that, due to the lack of access to one's actual audience, speakers target their contributions to an *imagined audience* (Marwick and boyd 2011). The imagined audience reflects how the writer wishes or expects to be interpreted. Whereas in the past, the task of imagining an audience was confined to authors of literature, it is now said to be a ubiquitous aspect of online communication that all online speakers engage in. Some philosophers have proposed developing a theory of "imagined common ground" to model online conversations (Lewis 2025) while others have suggested that the need for us to *imagine* our relationships with those with whom we communicate online plays an important role in online harassment (Barnes 2025). However valuable the notion of an imagined audience might be for helping us to understand the psychology of online language users, it doesn't appear to track our semantic intuitions about the reference of indexicals. When a poster uses the expression, 'you', they might have an imagined audience in mind but the second person pronoun doesn't refer to this imagined person. Instead, it is most naturally read as referring to the person who reads the post. In this sense, posts operate semantically like a billboard announcing that 'Jesus loves you' (discussed in the literature on context sensitivity, e.g. Egan 2009). Egan calls such posts *utterance bombs* as they are planted in one context and later detonated when opening in another context producing a new utterance (Egan 2009). The content of these expressions is determined, not only by the context in which they were produced but by the context in which they are received by users such that each person who engages with the post will be picked out by the second-person pronoun. By granting speakers the possibility of an unlimitedly wide audience while ensuring that their linguistic actions can have their effect long after the moment of their production, online communication vastly enhances a speaker's ability to plant such bombs.

The first-person pronoun raises its own distinct issues. For example, many platforms are populated by corporate accounts run by multiple individuals as well as anonymous and sockpuppet accounts (Paterson 2024).[5] While people may agree that, when a celebrity account run by an intern posts 'I like cheese', the 'I' does not necessarily denote the intern, it is less clear who it does refer to. It's intuitive to say that it refers to the celebrity in whose name the account is run, but this reading runs into challenges as it can be overridden by other contextual signals. The celebrity, call them Joe Blogs, may later post 'I would like to apologise for that last post. I have always opposed cheese jb'. If the signature serves to make out the reference as *genuine* in this case (or in any way different), then there is reason to believe that in the standard case, it does not have

[5] Anonymity raises distinctive legal and political concerns with countries varying on whether it falls under free speech protection. The debates on this issue reflect both how online speech is conceptualized differently from offline speech as well as how media ideology can, at least officially, vary across borders. In the US, *Reno v the ACLU* resulted in the ruling that there was no reason to qualify first amendment protections for online speech in a manner comparable to broadcast speech. Whereas the *ECHR KU v Finland* ruling held that freedom of expression may not cover anonymous advertisements. The UN special rapporteur for freedom of expression has explicitly supported to right to communicate anonymously online. See Barendt (2020, esp. ch.6) for discussion.

the same reference even if the apology suggests they are accepting responsibility for the post. There could be a range of explanations for this. It may be that the default usage is a kind of licensed fiction, denoting a fictional version of the celebrity, or that the text performs a vicarious speech act and denotes the real celebrity. The media ideology of a platform contains norms regulating and resolving the deferred reference involved in proxy posts but as these norms are in flux, intuitions about particular cases may vary.

The case of locative expressions shows that appeals to the context of interpretation don't help us to make sense of all demonstrative expressions. Consider, a user posting 'people on here are very uncharitable' or quote-posting another post and saying 'I hate it here'. The 'here' in these cases neither refers to the physical location of the poster which will be unknown to most other users nor does it function like an utterance bomb, picking out the location of the post's reader. Even if we set aside physical location and focus on the poster and the reader's own feeds or homepages as possible referents, the 'here' in the post doesn't pick out either of these. Different users, even if they follow exactly the same people, are unlikely to have the same feeds or access to the content of others'. Instead, the 'here' appears to denote either the platform or some proper part of it, e.g., 'philosophy Bluesky', 'linguistics Bluesky' – a region of 'cyberspace'. If this is the case, it may be worth altering the standard model of context for the analysis of online speech to include 'website' along with the time and physical location of the speaker.

Eliot Michaelson has argued that recording technologies (of which we can understand social media as a variety) introduce their own distinctive contexts with 'channel-dependent' metarules for semantic interpretation. For example, the first-person indexical used in a postcard denotes the writer, while for an answering machine it is more likely to pick out whoever *owns* the line (Michaelson 2013). These metarules are a manifestation of the channel's media ideology. We can see how different platforms might exhibit different metarules by comparing the Bluesky user posting 'I hate it here' with someone posting the same message on Facebook, a site that doesn't generate the same illusion of a common public sphere promoted by Bluesky.[6] In this instance, it is natural to interpret the 'here' as denoting the poster's own location rather than a shared imagined space. In general, the context determining the metarules for online communication appears to depend upon the platforms used. One aspect of the media ideology of the platform appears to be that its users conceptualize it as a digital location, somewhere which can be picked out by indexicals like 'here'. While this is not unique to Twitter (people may speak of a WhatsApp group as a location), it is not a ubiquitous feature of online communication and so it indicates an interaction between platform and ideology. Some platforms are conceptualized as public squares, others as noticeboards, others as semi-public journals.

---

[6] It is worth noting that Meta has ceased to describe Facebook and Instagram as social media sites functioning to connect people with friends and family. Just 20% of content people see on Facebook and 10% on Instagram comes from friends and family while the rest comes from influencers, media companies, and AI slop farms (Chayka 2025). How a social network markets itself to users and its actual business model may play independent and conflicting roles in the formation of media ideology.

With this in mind we can ask, is there anything unique about the notion of online contexts? If we focus exclusively on the issue of pragmatic impoverishment, then online communication is not distinctly novel. Uncertainty about who one's audience might be is a fundamental component of all written communication and so does not itself demonstrate the novelty of online speech. Telegrams and letter writing are both textual and asynchronous vehicles of communication. However, the public nature of social media platforms undermines both the traditional iterative structure of common ground and the capacity to establish stable norms of communication. There is also reason to believe that different platform infrastructures encourage users to imagine their context differently. Some networks encourage users to imagine themselves as communicating in a shared digital space in a way that can influence the interpretation of demonstrative expressions. I have suggested that the norms that underlie these processes of interpretation should be thought of as a part of the media ideology of the platform. In the next section, we will look at the form that traditionally non-verbal cues take on social media platforms.

28.2 Speech Acts

A second research program in the study of online speech has been the attempt to document and explain the new speech acts online communication enables. The idea that technological development facilitates the emergence of novel speech acts is not confined to online communication. Classic examples of speech acts facilitated by writing are the act of applying one's signature to a contract, drafting a law, or writing a medical prescription. A challenge for this line of research comes from the observation that the norms governing online discourse are in flux and different platform's media ideologies are yet to be decided.

At their core, these speech acts often involve some form of 'posting'. For the purposes of analysis, we will not treat posting as its own kind of speech act but as the act of adding any content to an online context; posting is to online speech as utterance is to verbal communication.[7] Unlike utterances, however, a post is not typically a response to another immediate conversational prompt. Posts often have the appearance of public announcements rather than contributions to a conversation (Goldberg 2021). We will confine ourselves here to the most discussed acts; sharing and 'liking' and set aside less studied acts like prayer (Öhman, Gorwa, and Floridi 2019), hash-tagging (Kukla 2024; Scott 2022), meme-ing (Ernst 2025), and @-ing (Goldberg 2021). These acts play an important role in the content-aggregation processes of media platforms, they exploit properties like asynchronicity and copyability, and they appear to convey distinct and possibly exclusively online contents. Again, our focus will be on the two structuring questions of this chapter; to what extent are these speech acts novel? And how does this novelty connect to the platform upon which these speech acts are taking place?

---

[7] There is little consensus on this point. Rini and Nguyen suggest that posts are (generally) assertions (Rini 2017; Nguyen 2021) and so users implicitly vouch for their truth. Record and Miller argue that posts themselves can express many different speech acts including directives and interrogatives (Record and Miller 2022), while Goldberg suggests that a post is a primitive speech act which is not subject to the same norms as assertion (Goldberg 2021). Browning and Adams hold that there are no constitutive rules for what a post involves (Browning and Adams 2023).

28.2.1 Sharing, Reposting

This section will look primarily at the category of sharing (e.g., retweeting) and for practical purposes, there will be a particular focus on Twitter (or X). There is relative consensus both that sharing is a uniquely online speech act which does not correspond to anything offline (Marsili 2021: 10461) and that a share is not itself an endorsement even if it may often be interpreted that way. The task of a theory of retweeting has been to characterize the act of retweeting, whether there is a common structure to all performances exhibiting a shared set of norms or communicative function, and to explain why the act is so often understood as a (defeasible) form of endorsement.[8] This latter demand is of particular importance as there is empirical evidence that people standardly interpret reposts as endorsements (Metaxas et al. 2017), which has further significance for the epistemic responsibilities of online agents in relation to the spread of 'fake news' and other disinformation. In what follows it will be helpful to distinguish bare retweets from quote-tweets and captured posts (i.e., in which the original post is screenshot either with or without the original poster's name redacted).[9]

To quickly dispense with the idea that retweets are always endorsements, it is observed that retweets are not confined to posts with assertoric or even linguistic contents. For example, the most retweeted post currently on Twitter is the Japanese billionaire Yusaku Maezawa's promise that he will give a million yen to 100 people who retweet his original post. Even if people who have retweeted this post have a positive attitude to its content, they shouldn't be understood as either reasserting it or even endorsing its truth-conditional content since the original post was a promise. One may not only reshare a post one disagrees with but instead to explicitly disagree with it as a quote-post (Habgood-Coote 2020). If not endorsement, then, what is the function of a retweet?

One prevalent view is that the function of retweeting is to call attention to the original post, regardless of whether its content is endorsed (Arielli 2020; Goldberg 2021:186; Marsili 2021; Record and Miller 2022). There are several ways this idea has been developed. While individuals may have different motivations for calling attention to a post, to share it is to implicitly at least, affirm its *shareworthiness*. Arielli takes this to be tantamount to an assertion of the original shareworthiness of the original post, "An act of sharing is therefore a speech act whose aim is to direct the attention of other people to a content, stating (or expressing) its shareworthiness" (Arielli 2020). In this version of the view, a retweet is simultaneously a directive, instructing others to pay attention to a post, and an assertion that the post is worth their attention.

---

[8] This, of course, assumes that there is a single category of retweeting. For example, Cappelen and Dever (2019) advocate pluralism about retweeting according to which it sometimes functions to reassert the original post, while at other times makes the meta-assertion that another has asserted the original post or merely promotes the original content. Frigerio and Tenchini argue that, while retweets can be used to perform communicative acts, they do not necessarily have any communicative significance and should be understood as a form of copying akin to photocopying a document (Frigerio and Tenchini 2023). Pepp, Michaelson, Sterken have also modified their earlier position to claim that sharing a post cannot be assigned a unique illocutionary significance (Pepp, Michaelson and Sterken 2025).

[9] Platforms like Mastodon do not allow quote posting, a deliberate choice on the designer's part "because it [quote-tweeting] inevitably adds toxicity to people's behaviours. You are tempted to quote when you should be replying, and so you speak at your audience instead of with the person you are talking to. It becomes performative" (Rochko 2018).

Accordingly, it would be incoherent to both share a post and say that it is not worth anyone's attention, e.g. 'I share p but do not have the intention that p is noticed'.

The attention approach faces several challenges. One problem is that a poster may perform the act of sharing with full awareness that everyone who follows them has already seen the original post (Frigerio and Tenchini 2023; Pepp, Michaelson, and Sterken 2025). If this is the case, it is argued, then the act of calling attention is not essential to sharing and so the act's illocutionary significance should be sought elsewhere. Intuitions about this case will likely depend upon whether one views uptake to be essential to the performance of a speech act, however, it is worth noting some responses the attention-based theorists could make. The act of assertion may still have the function of adding information to the common ground even if it is sometimes performed redundantly or with the knowledge that it will be unsuccessful. Similarly, one might think that the function of sharing is to call attention to a post even if it does not successfully fulfil this in all cases. Redundant sharing may even perform further speech acts in virtue of the fact that the function of sharing is to call attention. Sharing a post to followers who have already seen the original post violates Grice's maxim of quantity which demands that speakers be as informative as required. As a redundant retweet is not informative, it may serve secondary effects, e.g., signalling affiliation.[10] Alternatively, one might view redundant sharing as part of a group speech act. Just as each member of a booing crowd may believe their contribution to be superfluous for the crowd's overall ability to express opposition, this doesn't entail that the function of booing is not to express opposition.

Others deny that there is a propositional content implicit in the act of sharing, suggesting instead that it is more like pointing at the original post (Marsili 2021; Pepp, Michaelson and Sterken 2024). Marsili argues that retweets are a kind of *digital ostension*, a uniquely online act that shares properties with both traditional ostension and quotation. Like ostention, it makes manifest an intention to make something manifest, while like quotation, what is made manifest is a representation. We'll focus on the quotation aspect first. While it is not controversial to claim that one can quote non-verbal acts (e.g. 'he was like [speaker makes a grimace]') quotation has a further assertoric force in that it attributes a quoted statement to an individual. To quote a speaker is to endorse the claim *that* they have said something, and as such, it is to make a truth-evaluable assertion. What is distinctive about this account is the role played by the digital infrastructure. In the case of retweeting, the attributive component is built into the platform, anyone can see the original post, and so the assertoric force is removed from the act itself. In effect, the retweeter no longer asserts that someone said something, they show it. Further support for this claim can be found in how we interact with screenshots of posts. In some cases, to avoid generating direct attention (i.e., reposts) for an original post, individuals will screenshot the post and share that image rather than reposting the original. If one falsifies such an image, one will rightly be held responsible for the dishonesty involved. This suggests that retweeting carries with it the assertoric force of an attribution and so contains at least the implicit claim that the original post was made.

---

[10] It has been argued that in some cases the sharing of fake news stories is a matter of signalling group affiliation (Ganapini 2021) whereas others are critical of the idea that they serve a single function.

The second question concerns why a retweet is so often understood as an endorsement. There are several reasons why one might take retweeting to be a form of endorsement. Rini suggests that, while retweets are not themselves assertions, they are typically mistaken as assertions and thus this widespread misunderstanding underlies the endorsement interpretation (Rini 2017). In contrast to this mistaken-assertion view, others have suggested that distinctive norms are in place for sharing. Record and Miller suggest two competing norms for sharing news stories, a conditional norm which states that a story should be shared iff it would be interesting *if* it were true and a disjunctive norm which holds that a story should be shared iff it is *either* true or interesting. Empirical evidence suggests that both these norms are currently in operation (Altay, Hacquin, and Mercier 2020; Altay, Araujo, and Mercier 2021). If retweeting is governed by these norms, it may be reasonable for interpreters to infer that a person who retweets some content takes that content to be true. Alternatively, Marsili suggests that interpreters infer that the retweeter takes the original post to be true, not because of a norm particular to the speech act but in order to preserve the relevance of the original poster's contribution (where relevance is understood in the manner of Sperber and Wilson, 1995). What matters here is that the existence of multiple competing norms provides some evidence for the claim that media ideology, in particular, the norms governing how acts like retweets are interpreted, is not uniquely determined by the infrastructure of the platform upon which the acts are occurring.

28.2.2 'Liking'

The second most studied online communicative action is 'liking'. Historically, gestures such as 'likes' were introduced to address an aesthetic concern that Facebook designers had termed 'the redundant problem'(Morgans 2017). This problem arose from multiple people making the same comment under a post, for example, posting 'congratulations' after an announcement of an engagement. The influx of 'redundant' expressions could make it difficult for users to identify novel contributions to a discussion. Since the introduction of the 'like', Facebook has extended its repertoire of responses to include symbols for 'care', 'wow', 'sad', 'angry', and others. While it is unclear whether these should be understood as serving the same general kind of function, we will focus here on the traditional like which has been extended to platforms such as Twitter (now X), Bluesky, Mastodon, and others. For space reasons, we won't look at Reddit's upvoting (and downvoting) system which has its own distinctive norms.

At present there is a small but emerging consensus that 'likes' and other gestures serve phatic functions (Schandorf 2013). McDonald argues that 'likes' don't have a single, identifiable content, either propositional or emotional (McDonald 2021). On the one hand, we shouldn't take them to have propositional content since, if they did, one could raise questions about their truthfulness or the sincerity of the one who posted it, and these don't seem like options.[11] At the same time, they don't appear to express a consistent sentiment (Kukla 2024). They can be used both as markers of support and also to call attention to harmful political

[11] While 'likes' were advertised to users as equivalent to the phrase 'I like this', it's clear that their function has moved beyond indicating positive attitudes. The failure of the designers' intentions to fix the function of this communicative tool may encourage some caution in those who believe in technological fixes to how people use media platforms. For example, Sinan Aral has proposed introducing buttons for 'reliability', 'wholesomeness', or 'educational' indicating that the user takes the post to be reliable, etc. (Aral 2021). It isn't clear we should expect these designer-intended readings to survive in the hermeneutical maelstrom of the internet.

practices. Instead, McDonald argues, 'likes' function as backchannel responses akin to nods and other non-verbal cues that provide a speaker with 'positive face' (a similar view can be found in Lewis 2025) without entailing explicit endorsement.[12] While this account of 'liking' emphasizes the signal sent to the speaker, 'likes' also send signals to other conversational participants. For example, social media users are more likely to 'like' posts that have already been 'liked' (Sherman et al. 2016) and Alex Davies has argued that 'likes', by providing information about how others are reacting to a post, can alter how readers semantically interpret the posts contents (Davies 2025).

It is plausible then that 'liking' serves functions previously performed by non-verbal cues like smiling, nodding, and maintaining eye contact. Rather than introducing a new kind of expressive power into online communication, these online gestures serve to help users achieve communicative functions that would otherwise be carried out by non-verbal cues in offline communication. This interpretation of 'likes' makes them more like many emojis than assertions. Face and hand emojis, in particular, have been argued to serve the function of non-verbal cues that accompany speech (Gandiva et al. 2019; Gawne and McCulloch 2019).[13] However, unlike non-verbal cues, 'likes' are not ephemeral, they cling to a post long after the action of 'liking' has occurred. This enables them to be aggregated, to bring a post to prominence, and gives them a unique capacity to influence linguistic behaviour online.

The discussion so far has focused on the communicative significance of the actions that can be taken on online platforms. However, some researchers have argued that silence, too, takes on different meanings in online communication. Alexander Brown argues that silence on social media does not have the same significance as silence offline (Brown 2019). This may require us to rethink arguments to the effect that silence licenses (McGowan 2012), lends authority to (Maitra 2012), or default acceptance to (Goldberg 2021) hate speech in online contexts. One upshot of these views is that theories of counter-speech which were developed for offline conversations may be less applicable to online ones (Lackey 2021; Saul 2021). Goldberg suggests that, if non-verbal indicators of assent and dissent must be replaced by explicit verbal acts online, then people may feel the need to comment on an issue online that they wouldn't otherwise comment upon offline on the grounds that failure to comment may appear to signal indifference, rather than a lack of something useful to contribute (Goldberg 2021). In the next section, we will look more at how features of online communication infrastructure introduce unique incentives for linguistic behaviour.

Returning to our two framing questions, we may ask if these speech acts are genuinely novel and to what extent their features depend on the infrastructure of communication. Kukla and Marsili both claim that online communication facilitates novel speech acts that cannot be identified directly with offline analogues. According to Marsili "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

---

[12] The question of whether 'likes' are endorsements is legally significant. In 2020, a Zurich court ruled that a person could be charged with defamation for 'liking' a social media post on the grounds that 'liking' indicated endorsement and constituted an act of dissemination.

[13] Though see Grosz for the claim that emojis function as expressions like 'wow' (Grosz et al. 2023), Maier and Albert for the claim that emojis function as pictures (Maier 2023; Albert 2020) and McDonald and King who both compare emojis to artistic performances (King 2017; McDonald 2024).

share it or *retweet* it" (Marsili 2021: 10459), while Kukla writes that "technologically enabled online speech acts… have distinctive pragmatic structures, which are irreducible to speech-act types performed over other media or in person" (Kukla 2024). In contrast, more conservative interpretations hold that retweeting and 'liking' accomplish nothing that can't be accomplished by non-verbal cues like pointing or smiling. Even if they provide us with new ways of accomplishing familiar communicative acts, they do not enhance our expressive capacities. The fact that our norms for interpreting these actions are yet to be fully settled makes it difficult to assess whether the conservative or revolutionary positions are correct.

28.3 The Impact of Platform Infrastructure on Discursive Goals

Over the past decade, a subfield of books and articles have posited that social media plays a role in the decline of epistemic norms (Rini 2017), the exacerbation of political polarization (Settle 2018; Sunstein, 2018; Seymour 2019; Bail, 2022), and the increased fragility of political organization (Tufecki 2017). One of the most influential examples supporting the idea that platform infrastructure influences linguistic behavior is Nguyen's argument that Twitter's technological structure transforms the inherent goals of discourse through the process of gamification (Nguyen 2021).

According to Nguyen, Twitter alters the goals of conversation through *value capture*, a process that occurs when rich, subtle, and hard-to-express natural values are placed in a social or institutional setting that presents simplified and typically quantitative metrics. These simplified and quantified measures then become the focus of individuals' motivations and decision-making processes. The classic example of this is when a person begins to exercise for the sake of their physical and mental well-being but over time comes to prioritize increasing the step count on their FitBit instead. In the case of Twitter, discursive practices which normally have truth, knowledge, or mutual understanding as their goals, are captured by the metrics of 'likes', retweets, and follower counts so that, through engaging with Twitter, agents have their linguistic goals and ultimately their linguistic behaviour altered as a result of the new metric. According to Nguyen, "Twitter shapes our goals for discourse by making conversation something like a game. Twitter scores our conversation" (Nguyen 2021 — see also McDonald 2021; Lewiński and Dutilh Novaes 2025).

Nguyen's work provides one of the clearest case studies of how particular features of an online platform can influence linguistic behaviour. It is crucial to be clear about what the underlying mechanism at play.[14] Twitter makes salient the numerical representation of a post's 'likes' and retweets, as well as a user's number of followers. This simplified score facilitates hedonic instrumentalization by presenting users with a target for their online activity. The pleasure of pursuing this target can, for some users, become the object of their motivations, at the expense of richer more complex values such as the pursuit of truth. Furthermore, certain

---

[14] For example, Browning and Adams suggest that Nguyen's account relies on a strict analogy between Twitter and games according to which Twitter must have constitutive rules that sculpt the agency of users and "because the service lacks constitutive rules, it does not sculpt agency in any interesting way" (Browning and Adams 2023). They further suggest that "Nguyen is proposing that Twitter has built-in norms that shape how we are supposed to use the technology" (Browning and Adams 2023). I think both these readings are too strong. The concept of a norm plays no role in the gamification model while Nguyen distinguishes between the few players for whom Twitter is explicitly a game (i.e., game-playing users), and those who Twitter merely incentivizes to modify their behaviour. Users of Twitter may not conceive of themselves to be playing a game for the hedonic incentive structure described by Nguyen to influence their behaviour.

methods of engagement are more likely to increase these numerical values than others, for example, statements of moral outrage (Brady et al., 2021).[15] Over time, users come to prioritize the pursuit of simple pleasures over complex values. Some critics of the gamification model have objected to the claim that users are not pursuing 'likes' and retweets per se rather what those 'likes' and retweets represent, e.g., social esteem (Williams 2022). In this sense, online communication is no different from offline communication where our behaviours are driven by similar incentives of group-membership, esteem, and other social pressures.

Regardless of whether the process of gamification is a unique consequence of the communication bottleneck of social media in which rich non-verbal cues are compressed into discrete, enduring and quantifiable 'likes', Nguyen is correct that the owners of social media platforms have strong incentives to drive engagement and have unique abilities to manipulate user behaviour, a point made more salient by the purchase of Twitter by Elon Musk in 2022. The incorporation of chatbots into online discourse (Connolly 2025) and the potential role of these chatbots as stochastic stool pigeons 'aligned' to espouse the positions of their owners present further risks.

28.4 Conclusion

In this chapter, we have looked at several features of online speech that are claimed to be linguistically unique. We have seen that in most cases, the supposed novelty of online speech is a matter of degree rather than a difference in kind. Pragmatic impoverishment, context collapse, utterance bombs, are all familiar features of written communication. Similarly, it appears that many of the unique speech acts facilitated by online communication are attempts to reconstitute in the digital domain familiar features of offline communication. Sharing and liking manage joint attention in the absence of non-verbal cues. Nevertheless, there is also evidence that these features give rise to properties not found online. In particular, the textual nature of online communication ensures that a non-verbal signal, for example, a 'like', that would be ephemeral in offline discourse endures in the online context in a manner which allows for its significance to be altered. If this is correct, this property generates distinctive incentives for online speech which, in turn, influence linguistic behaviour. It may then be the case that, while the online context doesn't radically alter what our language is, it does alter how we use it.

---

[15] Specifically, the authors found that the presence of moral-emotional terms increased the diffusion of a post by 20% for each moral term. There is additional evidence that falsehoods travel further and faster than truths on social media (Vosoughi, et al. 2018) There is some empirical evidence to support the claim that social media engagement utilises the dopaminergic system behind reinforcement learning (Lindström et al. 2021). However, Nguyen's account doesn't appeal to this particular mechanism, merely that users find pleasure in increasing their 'likes', retweets, and follower counts. For more detail about how a social media site can serve as an affective or attentional scaffold, see Kreuger and Osler 2019 and Voinea, Marin, and Vică 2024.

Bibliography[16]

Albert, Georg. 2020. “Beyond the Binary: Emoji as a Challenge to the Image-Word Distinction”. In *Visualizing Digital Discourse: Interactional, Institutional and Ideological Perspectives*, eds. C. Thurlow, C. Dürscheid and F. Diémoz, 65–80. Berlin: De Gruyter Mouton.

Altay, S., de Araujo, E., & Mercier, H. 2021. “‘If This account is True, It is Most Enormously Wonderful’: Interestingness-if-true and the sharing of true and false news”. *Digital Journalism*, https://doi.org/10.1080/21670811.2021.1941163

Altay, S., Hacquin, A-S, and Mercier, H. 2020. Why Do so Few People Share Fake News? It Hurts their Reputation. *New Media and Society,* https://doi.org/10.1177/1461444820969893

Aral, S. 2021.The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health—and How We Must Adapt. Currency

Arielli, E. 2018. “Sharing as speech act.” *Versus*, 127(2):2, 243–258.

Bail, C. 2022. *Breaking the social media prism: How to Make Our Platforms Less Polarizing*. Princeton University Press.

Barnes, M. 2025. “Complex Harms in Online Speech: The Limits of the Illocutionary” In *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.

Beaver, D. I., C. Roberts, M. Simons, and J. Tonhauser. 2017. “Questions under Discussion: Where Information Structure Meets Projective Content. *Annual Review of Linguistics*, 3:265-284.

Benton, P., & Schmidt, M. W. 2024. “The Harm of Social Media to Public Reason”. *Topoi*, *43*(5), 1433-1449.

Bergamaschi Ganapini, M. 2023. “The signaling function of sharing fake stories”. Mind & Language, 38(1):64-80.

Bernstein, M., A. Monroy-Hernández, D. Harry, P. André, K. Panovich, and G. Vargas. 2021. 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *Proceedings of the International AAAI Conference on Web and Social Media*, *5*:50-57. 1.

Bianchi, C. 2014. How to do things with (recorded) words. *Philosophical Studies* 167:485–495. https://doi.org/10.1007/s11098-013-0111-0

Brady, W. J., K. McLoughlin, T.N. Doan, and M.J. Crockett. 2021. How social learning amplifies moral outrage expression in online social networks. *Science Advances*, 7(33):e, eabe5641.

Brown, A. 2019. “The Meaning of Silence in Cyberspace: The Authority Problem and Online Hate Speech” In *Free Speech in the Digital Age*. Oxford University Press.

Browning J. & Adams Z. 2023. Why Twitter does not gamify communication, *Inquiry*, 1-18 DOI: 10.1080/0020174X.2023.2261489

Camp, E. 2018. “Insinuation, common ground, and the conversational record” In New Work on Speech Acts, edited by D. Fogal, D. W. Harris, and M. Moss. Oxford University Press.

Chayka, K. 2025. “Mark Zuckerberg Says Social Media Is Over” *Vanity Fair*, Available at: https://www.vanityfair.com/news/story/mark-zuckerberg-is-so-over-facebook-friends?

---

[16] Thanks to Austin Baker and Gabe Dupre for comments on earlier versions of this.

srsltid=AfmBOoo6neVvoxwK-ycNfZ6Aoe3csXRNXU8j8OZgf58Y0bsdbEl-E48L (Accessed 7th August 2025)

Connolly, P. 2025 “Conversations with Chatbots” in *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.

Davis, J. L. 2020. *How Artifacts Afford: The Power and Politics of Everyday Things*. MIT Press.

Egan, A. 2009. “Billboards, Bombs, and shotgun weddings” *Synthese*, 166:251–279.

Eisenstein, E. 1979. *The Printing Press as an Agent of Change: Communications and Cultural Transformations in Early Modern Europe.* Cambridge University Press

Ernst, N. 2025. “A Meme for Excuses” In *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.

Frigerio, A., and M. P. Tenchini. 2023. “Sharing.” Ratio 36:147–158. doi:10. 1111/rati.12375.

Frost-Arnold, K. 2021. “The Epistemic Dangers of Context Collapse Online.” In *Applied Epistemology*, edited by J. Lackey, 437–456. Oxford University Press.

Gawne, L, & McCulloch G. 2019. “Emoji as Digital Gestures”. *Language@Internet* 17(2). https://www.languageatinternet.org/articles/2019/gawne.

Gershion, I. 2010. “Media Ideologies: An Introduction”. *Journal of Linguistic Anthropology* 20(2):2 283–293.

Goldberg, S. C. 2021. “The Promise and Pitfalls of Online ‘Conversations’.” Royal Institute of Philosophy Supplements 89:177–193

Grosz, P. G., G. Greenberg, C. De Leon, and E. Kaiser. 2023. “A Semantics of Face Emoji in Discourse.” *Linguistics and Philosophy*, *46*(4), 905-957.

Habgood-Coote, J. 2020. Retweets, Endorsements, and Indirect Speech Acts. Retrieved from https://www.prindleinstitute.org/2020/07/retweets-endorsements-and-indirect-speech-acts/.

Haiman, J. 1998. *Talk is Cheap: Sarcasm, Alienation, and the Evolution of Language*. Oxford University Press.

Hoek, D. 2018. “Conversational Exculpature.” *Philosophical Review*, 127(2):151

Kaplan, D. 1979. “On the logic of demonstratives.” *Journal of Philosophical Logic* 8:81-98.

King, A. 2017. ‘A Plea for Emoji’. In American Society for Aesthetics Newsletter, edited by Shelby Moser and Michel-Antoine Xhignesse, 1–3.

Krueger, J., and L. Osler. 2019. “Engineering Affect.” *Philosophical Topics*, 47(2):205-232.

Kukla Q. 2024. “Don’t @ Me! The Pragmatics of Online Speech Acts” In the *Oxford Handbook of Applied Philosophy of Language*, edited by Luvell Anderson and Ernest Lepore.

Lackey, J. 2021. “Echo Chambers, Fake News, and Social Epistemology.” In *The Epistemology of Fake News*, 206-227. Oxford University Press.

Lewiński, M., and C.D. Novaes. 2025. “The Many-to-Many Model: Communication, Attention, and Trust in Online Conversations.” In *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.

Lewis, D. K. 1979. “Scorekeeping in a Language Game.” *Philosophical Logic*, 8(1):339– 359.

Lewis, K. 2025. “Imagined Audiences and Common Ground” In *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.

Lindström, B., M. Bellander, D. T. Schultner, et al. 2021. A Computational Reward Learning Account of Social Media Engagement. *Nature Communications* 12:1311 https://doi.org/10.1038/s41467-020-19607-x

Maier, E. 2023. “Emojis as pictures.” *Ergo*, 10(12):317-355.

Maitra, I. 2012. “Subordinating Speech.” In *Speech and harm: Controversies over free speech*, 94-120 Oxford University Press

Maitra, I. 2025. “Subordinating Speech Online, or Why RTs *are* Endorsements” In *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.

Marsili, N. 2021. “Retweeting: Its Linguistic and Epistemic Value.” *Synthese*, 198(11):10457–10483.

Marwick, A. E., and d. boyd. 2011. “I Tweet Honestly, I Tweet Passionately: Twitter users, context collapse, and the imagined audience.” *New Media & Society*, 113(1):114–133.

McCulloch, G. 2020. *Because Internet: Understanding the New Rules of Language.* Penguin.

McDonald, L. 2021. “Please Like This Paper.” *Philosophy*, *96*(3):335-358.

McDonald, L. 2024. “Making Sense of Emoji.” *Philosophy*, *99*(3):413-435.

McDonald, L. 2025. “Context Collapse Online” In *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.

McGowan, M. K. 2012. “On ‘Whites Only’ Signs and Racist Hate Speech: Verbal Acts of Racial Discrimination.” In *Speech and Harm: Controversies over Free Speech*, 121–147. Oxford University Press.

McLuhan, M. 2017. “The Medium is the Message.” In *Communication Theory*, 390–402. Routledge.

McMullan A. .2025. “Sarcasm in Online Discourse” In *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.

Miller, B., and I. Record. 2013. “Justified Belief in a Digital Age: on the Epistemic Implications of Secret Internet Technologies.” *Episteme* 10(2):117–134.

Miller, B., and I. Record. 2017. “Responsible epistemic technologies: A social-epistemological analysis of autocompleted web search.” *New Media & Society* 19(12):1945–1963.

Metaxas, P. T., & Twittertrails Research Team TTRT. 2017. "Retweets Indicate Agreement, Endorsement, Trust: A Meta-Analysis of Published Twitter Research." *Arxiv*. Retrieved from http:cs.wellesley.edu/~pmetaxas/WorkingPapers/Retweet-meaning.pdf

Morgans, J. 2017. “The Inventor of the ‘Like’ Button Wants You to Stop Worrying About Likes.” *Vice.* Available at: https://www.vice.com/en/article/the-inventor-of-the-like-button-wants-you-to-stop-worrying-about-likes/ (Accessed 25 August, 2025)

Nguyen, C. Thi. 2020. “Echo chambers and epistemic bubbles.” *Episteme* 17(2):141- 161.

Nguyen, C. T. 2021. “How Twitter gamifies communication.” *Applied Epistemology* 5(6):410-436.

Ohman, C., R. Gorwa, and L. Floridi. 2019. “Prayer-bots and Religious Worship on Twitter: A Call for a Wider Research Agenda.” *Minds & Machines* 29(2):331-338.

O'Neill, O. 2022. *A Philosopher Looks at Digital Communication.* Vol. 4. Cambridge University Press.

Ong, W. J., & Hartley, J. 2013. *Orality and Literacy*. Routledge.

Pepp, J., E. Michaelson, and R. K. Sterken. 2024. “On Retweeting.” In *The Oxford Handbook of Applied Philosophy of Language*, edited by L. Anderson and E. Lepore. Oxford University Press

Pollock, J. 2024. “Epistemic Bubbles and Contextual Discordance.” *Philosophy*, 99(3):437-459.

Record, I., and B. Miller. 2022. *People, Posts, and Platforms: Reducing the Spread of Online Toxicity by Contextualizing Content and Setting Norms*. Manuscript in preparation

Rini, R. 2017. “Fake News and Partisan Epistemology.” *Kennedy Institute of Ethics Journal* 27(2S):E-43–E-64.

Rini, R. 2025. “Context Collapse and Pop-Up Communities.” In *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.

Roberts, C. 2018. “Speech Acts in Discourse Context.” In *New Work on Speech Acts* edited by D. Fogal, D. Harris, M. Moss, 317-59. Oxford University Press

Roberts, C. 2012. “Information Structure: Towards an Integrated Formal Theory of Pragmatics.” *Semantics and Pragmatics* 5:1–69.

Rochko E. 2018. 10 May. Available at https://mastodon.social/@Gargron/99662106175542726 (Accessed: 25 August 2025)

Rothwell, J., 2023. How parenting and self-control mediate the link between social media use and youth mental health. *Institute for Family Studies and Gallup*.

Saul, J. 2018. “Dogwhistles, Political Manipulation, and Philosophy of Language.” In *New Work on Speech Acts*, edited by D. Fogal, D. W. Harris, and M. Moss, 360–383. Oxford University Press

Saul, J. 2021. “Someone Is Wrong on the Internet: Is There an Obligation to Correct False and Oppressive Speech on Social Media?” In *The epistemology of Deceit in a Postdigital Era: Dupery by Design,* edited by A. MacKenzie, J. Rose, I. and Bhatt, 139-157. Springer Nature.

Schaffer, J., and Z. G. Szabó. 2014. “Epistemic Comparativism: A Contextualist Semantics for Knowledge Ascriptions.” Philosophical Studies 168:491–543

Schandorf, M. 2013. “ Mediated gesture: Paralinguistic communication and phatic text.” *Convergence*, 19(3):319-344.

Schoubye, A. J., and A. Stokke. 2016. “What is said?” *Noûs*, 50(4):759-793.

Scott, K. 2022. *Pragmatics Online.* Routledge.

Settle, J. E. 2018. *Frenemies: How Social Media Polarizes America*. Cambridge University Press.

Seymour, R. 2019. *The Twittering Machine*. The Indigo Press.

Sperber, Dan, and Deirdre Wilson. 1995. *Relevance Theory.* 2nd ed. Blackwell Publishing.

Stalnaker, R. 1978. “Assertion.” In Syntax and Semantics: Pragmatics 9, edited by P. Cole, 315–322. Academic Press

Sunstein, C. 2018. *Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.

Tanesini, A. 2022. “Affective Polarisation and Emotional Distortions on Social Media.” *Royal Institute of Philosophy Supplements*, 92:87-109.

Tanesini, A. 2025. “Commitment online.” In *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.

Tirrell, L. 2025. “Hit-and-Run Speech Acts.” In *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.

Tufekci, Z. 2017. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press.

Voinea, C., L. Marin, and C. Vică. 2024. “Digital slot machines: social media platforms as attentional scaffolds”. *Topoi*, *43*(3):685-695.

Vosoughi, S., Roy, D., & Aral, S. 2018. “The Spread of True and False News Online.” *Science* 359(6380):1146-1151.

Wagner, B. E., A. L. Folk, S. L. Hahn, D. J. Barr-Anderson, N. Larson, and D. Neumark-Sztainer. 2021. “Recreational Screen Time Behaviors during the COVID-19 Pandemic in the U.S.: A Mixed-Methods Study among a Diverse Population-Based Sample of Emerging Adults.” International Journal of Environmental Research and Public Health 18(9):4613. doi:10 . 3390/ijerph18094613.

Williams, D. 2022. “Identity-Defining Beliefs on Social Media.” *Philosophical Topics*, *50*(2):41-64.

Woodard, E. & Neufeld E. 2025. “On Subtweeting” In *Conversations Online* edited by P. Connolly, S. C. Goldberg, and J. Saul. Oxford University Press.