# On Subtweeting*

Eleonore Neufeld[1] and Elise Woodard[†2]

[1]Department of Philosophy, University of Massachusetts Amherst
[2]Department of Linguistics & Philosophy, Massachusetts Institute of Technology

## 1 Introduction

> both speech and writing are risky: speech is imprecise, with no opportunity for editing, but it's also ephemeral; with writing, your bad judgment is set in stone and easily circulated. luckily there's a form of communication that combines the worst of both: tweeting
>
> *@kukukadoo*

Philosophers have recently begun exploring the moral, epistemic, and linguistic implications of social media engagement. For example, several philosophers have begun to explore the communicative and epistemic profile of retweets (Rini 2017; Marsili 2021; Nguyen 2021b). Thi Nguyen has recently argued that one problem with Twitter is that it gamifies communication (Nguyen 2021a). And there is a growing literature on the epistemic dynamics of fake news and the roles that social media companies have in stopping its spread (Fritts and Cabrera 2022a,b; Habgood-Coote 2019).

Another area of research that has received considerable attention in recent years is *strategic speech*. Speech is generally risky: it can offend, upset, embarrass, and come with interpersonal, reputational, financial, or legal costs. Insinuation (Camp 2018; Fricker 2012), dogwhistles (Saul 2018; Guercio and Caso 2022), misleading (Saul 2012; Viebahn 2021; Marsili and Löhr forthcoming; Saul 2012; Wiegmann et al. 2022), figleaves (Saul 2017; Bräuer 2023), code words (Khoo 2017), and more have all been put forward as types of speech in which a speaker strategically makes use of certain conversational 'tricks' in order to minimize these risks.

Despite the philosophical interest in both strategic speech and social media engagement, there remains a type of speech act that hasn't yet gained attention in either of these domains:

*Acknowledgements: We thank Patrick Connolly, Guillermo Del Pinal, Sandy Goldberg, Joshua Habgood-Coote, Michael Hannon, Benj Hellie, Kat Hintikka, Maegan Fairchild, Renée Jorgensen, Nicola Kemp, Nick Laskowski, Junhyo Lee, Neri Marsili, Christa Peterson, Joey Pollock, and Jennifer Saul for helpful feedback and discussion. Thanks also to audiences at the 2021 Michigan-MIT Social Philosophy Workshop, Cal State Long Beach, the Words Workshop, the 2022 Pacific APA, and the Online Conversations Workshop.

†Since writing this article, Twitter has been renamed to "X". Throughout the paper, we'll still use "Twitter" and its variations ("retweeting", "subtweeting", etc.) to refer to the platform and conventionalized platform-specific activities.

*subtweeting*. Broadly speaking, subtweets occur when one Twitter user critically or mockingly tweets about a person without mentioning their name or username. Subtweeting is one of Twitter's paradigmatic forms of strategic speech. Moreover, it is a *systematic* phenomenon—so systematic that it received its own name. This chapter aims to elucidate the strategic aims of subtweeting and the mechanics through which it achieves them. We thereby hope to shed light on the distinctive communicative and moral texture of subtweeting while filling in a gap in the philosophical literature on strategic speech in social media.

The structure of this paper is as follows. In §2, we specify what subtweets are and identify the central features that give rise to its strategic mechanics. In §3 and §4, we draw attention to problematic aspects of subtweeting and consider conditions under which subtweeting can be justified on moral and prudential grounds. §5 concludes by discussing the practical upshots of this paper and noting avenues for future work.

## 2 What is Subtweeting?

### 2.1 *Subtweets, Risks, and Deniability*

On July 17, 2016, Kim Kardashian took to Twitter to publish the following tweet:

(1) @KimKardashian: "Wait it's legit National Snake Day?!?!?They have holidays for everybody, I mean everything these days!"[1]

The tweeting took place in the midst of a much-publicized feud between her and the singer Taylor Swift, just days before Kardashian would dedicate a substantial portion of a new episode of *Keeping Up With the Kardashians* to bad-mouth Taylor Swift explicitly. Notably, although Taylor Swift is nowhere mentioned in the tweet, everyone with the requisite background knowledge knew that Kardashian's tweet was about Taylor Swift.

Kim Kardashian's tweet is a paradigmatic example of a *subtweet*: a tweet in which one Twitter user critically or mockingly tweets about another Twitter user without mentioning either their username or their name.[2] At the same time, the subject of the subtweet can often still recognize herself as the target. Sometimes, the audience can reasonably infer the target as well, provided they have the relevant background knowledge—as in Kardashian's subtweet of Taylor Swift. In other cases, it isn't as easy to isolate a unique target of a tweet. Consider a fictional tweet by a senior professor of philosophy:

1. https://twitter.com/kimkardashian/status/754818471465287680. The text portion of the tweet was followed by several snake emojis.

2. According to one definition on Urban Dictionary, when we subtweet, we are "[i]ndirectly tweeting something about someone without mentioning their name. Even though their name is not mentioned, it is clear who the person tweeting is referring to." See https://www.urbandictionary.com/define.php?term=subtweeting. The Merriam Webster dictionary characterizes them as "a usually mocking or critical tweet that alludes to another Twitter user without including a link to the user's account and often without directly mentioning the user's name" (https://www.merriam-webster.com/dictionary/subtweet).

(2) @Prof_Leslie: Just saw an um... interesting talk on feminist philosophy. Glad I didn't give that talk!

Suppose that Leslie was targeting Heidi, a junior scholar whom Leslie has disliked ever since Heidi got the job that Leslie competed for. Suppose further that there are multiple philosophers who recently gave talks on feminist philosophy. While it will be clear to many that the tweet has a target, it will be unclear to many users whom the specific target is. At most, only a constrained subset of the audience—e.g., those who know about Leslie's distaste for the talk and its author—will be able to directly identify who the tweet is about.

Although subtweets come in many different forms, they are united by a number of common features. Let's start with some obvious ones. First, subtweets have some intended *target* or *subtweetee*: the person the subtweet is supposed to be, in some way, about.[3] Second, others can sometimes identify the subtweetee if they have the relevant contextual information, which may or may not be publicly available. Finally, and most importantly for our purposes, the subtweet is about a specific target only *indirectly*.[4] While other aspects of a subtweet's content are often left somewhat inexplicit too (e.g., Kim Kardashian does not say that the target of her tweet *is* snake-like), a key feature of subtweets is that their *target* is intentionally inexplicit.[5]

If one of the main functions of subtweets is to say something *about* a specific target, why would the author of a subtweet choose to get at their target only indirectly? One answer lies in the distinctively *strategic* aspect of subtweeting. The communicated content of subtweets generally says something offensive, critical, mocking, or otherwise risky about a target. When the author of a tweet intentionally crafts their message in a way that's indirect, they are able to mitigate that risk. This is the essentially strategic function of subtweeting. This function is characteristically *realized* via the plausible deniability that's generated by the tweet's indirectness.[6] Later on, we will qualify this statement slightly: in rare cases, the indirectness of subtweets can be used to afford subtweeters other ways of risk mitigation.

3. Note that it is possible that a subtweet's target is more than one person. For simplicity, we use the singular formulation throughout. An interesting question that we won't address here is whether it's possible to subtweet, say, institutions, coorporations, or organizations without subtweeting any person.

4. Edwards and Harris (2016) make this point even more strongly: "Subtweets are, by definition, indirect messages" (p. 306).

5. There is some debate about whether subtweets that mention someone's name explicitly but without "@-ing" them (i.e. using their Twitter handle) can count as subtweets (see e.g., Reinsberg (2013)). One definition on Urban Dictionary claims they can, while other sites (and most definitions on Urban Dictionary) disagree. (See, for example, `https://sova.pitt.edu/the-art-and-harm-of-subtweeting`.) Bogost (2015) helpfully dubs them "supertweets." We set these cases aside for now but return to them briefly in §3.

6. The phenomenon of plausible deniability has been investigated intensely by philosophers of language (Dinges and Zakkou forthcoming; Mazzarella 2021; Camp 2018; Berstler 2019; Peet 2015; Lee and Pinker 2010). It's still an open question how to best characterize it. Note that "plausible deniability" is a technical term in the general literature on strategic speech that describes the speaker's goal in risk mitigation techniques, and is often used interchangeably with 'deniability.' It is often acknowledged that having plausible deniability is, paradoxically, compatible with cases of actual denial being *implausible*. This will also apply to cases of subtweeting. For discussions of 'implausible plausible deniability' and the epistemic limits of plausible deniability, see Dinges and Zakkou (forthcoming); Camp (2018); Berstler (2019); Lee and Pinker (2010).

The mechanism through which the author of a subtweet achieves plausible deniability is analogous to other cases of plausible deniability, which often occurs via implicature generation. Hence, we should first say something about how implicatures work in general and then how this mechanism is exploited in cases of deniability. When a speaker $S$ utters $U$ with conventional content $P$ in context $C$ and implicated speaker meaning $M$, they intend hearer $H$ to use mutually salient presuppositions and interpretive assumptions $I$ to derive $M$ from the fact that $S$ uttered $U$ in $C$. For example, when a speaker utters, "He arrives to seminars on time" in response to a question about a student's philosophical abilities, it is clear to the hearer—given shared assumptions—that the speaker does not think highly of the student's philosophical talents. Cases of plausible deniability exploit the fact that, although the set of salient presuppositions $I$ is in fact obvious to the speaker and hearer, they're not *acknowledged* to be mutually obvious to both. Thus, when $S$ is challenged, they can then pretend:

> to be in a slightly different context $C'$, governed by an alternative set of interpretive assumptions $I'$, which differ from $I$ in relatively crucial but intangible ways, such as the relative ranking of salience among features or objects, or the relative probabilities of various counterfactual possibilities. Given these differences, the calculation of $U$ plus $I'$ delivers [$M'$ rather than $M$, where it's possible that $M' = P$] as $U$'s implicated content. (Camp 2018, p. 49–50)

The speaker who exploits deniability claims to have meant to convey something else, other than what has actually been implicated. Similarly, by alluding to their target only indirectly, the lack of explicitness endows the author of the subtweet with the ability to deny that their intended message was about a given target $T$. Instead, they can insist that they were in a different context set $C'$ with presuppositions $I'$ that wouldn't render $T$ as the target of the message.[7] This is true even in cases in which the target seems to be obviously identifiable, just not mutually acknowledged as such.[8] In line with this mechanism, we adopt this commonly-accepted characterization of deniability for the purposes of this paper:[9]

**Deniability** $S$ has deniability relative to $M$ iff it is reasonable to calculate that the speaker meant to convey $M'$ on the basis of the uttered sentence's conventional meaning $P$, the commitments $K$ undertaken in the conversation to this point, and some set $I'$ of epistemically accessible presuppositions consistent with those commitments, in a way that renders $S$'s utterance $U$ at least minimally conversationally cooperative.

7. Note that it is possible, both in cases of subtweeting and in cases of conversational implicature more generally, that $H$ isn't able to compute $M$ or $T$ while they are able to recognize *that* something or other has been implicated, or that there is some target or other. We will say more about this in §2.2.

8. Media scholars Alice Marwick and danah boyd also highlight plausible deniability as one of the main motivations behind subtweeting, noting that "[s]ubtweeting creates plausible deniability, since the subtweeter can always claim the tweet was about someone else if confronted" (Marwick and boyd 2014, 1059).

9. Cf. Camp (2018), p. 50, whose characterization and notation we (roughly) adopt here. See also Peet (2015) and Fricker (2012), though see Dinges and Zakkou (forthcoming) for criticism.

In order to give an intuitive illustration of how subtweets' indirectness helps the speaker achieve deniability, consider again our hypothetical subtweet (2), reproduced below:

(2) @Prof_Leslie: Just saw an um... interesting talk on feminist philosophy. Glad I didn't give that talk!

Suppose that Leslie is criticized for targeting Heidi. Given the inspecifity of the target, Leslie can deny that the subtweet was about Heidi but rather someone else. In fact, Leslie can deny that she intended to say something negative about any feminist philosopher. If someone reproaches Leslie not for criticizing Anastasjia, but for publicly criticizing a feminist philosopher, Leslie can insist that she merely meant that she was glad to not have to deal with the combative questioners or to have experienced some technical difficulty; she merely didn't mention the identity of the speaker since it wasn't relevant to her conversational point. In each case, the possible deniability arises from the tweet's indirectness.

### 2.2 *Subtweeting and Other Forms of Strategic Speech*

As mentioned earlier, plausible deniability has often been emphasized as a key aspect of many other types of strategic speech. We'll briefly review the phenomena of insinuation, misleading, and dogwhistles before pointing to similarities and differences between them and subtweeting. As we will see, this contrastive exercise reveals important lessons for the distinctive strategic texture of subtweeting.

Let's start with dogwhistles. Dogwhistles are commonly described as speech acts that allow for two plausible interpretations: the first, innocuous interpretation is targeted toward a general audience or 'outgroup,' while the second, objectionable interpretation is targeted to a subset of this audience, the 'ingroup.' Crucially, the latter is concealed such that the general audience is unaware of the second, coded interpretation (Henderson and McCready 2018; Witten 2014; Saul 2018). In the United States, for example, 'inner city' has become a dogwhistle for 'Black' (Saul 2018). When the locution is used by politicians in constructions such as 'inner city crime,' it can activate pre-existing racial resentments of certain voters while preserving deniability about violating the 'Norm of Racial Equality' (Haney-López 2014; Saul 2018).[10] For example, imagine a politician who vows to crack down on 'inner city crime.' If criticized for racist speech, the politician might deny the charge, exclaiming: "What? I only talked about crime in inner cities! You are the one who is racist and thinks this has anything to do with race!" One of the central points of dog whistling is thus to get certain contents across while preserving deniability and avoiding liability if met with pushback.[11]

10. The 'Norm of Racial Equality' is a term coined by Tali Mendelberg (2017), who uses it to describe a socio-normative shift in American political discourse, happening between 1930 and 1960, concerning the permissibility of overtly and explicitly expressing racist attitudes. See also Saul (2018).

11. A detailed discussion of dogwhistles would require us to consider the spectrum of nuances related to dog-whistles, such as the difference (if any) between overt and covert dogwhistles, and intentional and unintentional

Next, consider the practice of *misleading*. The classical characterization of misleading claims that, in contrast to lying, misleading implicates but doesn't *say* something the speaker believes to be false (Berstler 2019; Viebahn 2021; Marsili and Löhr forthcoming). Consider this well-known example for illustration (from Saul 2012):

> A dying woman asks a doctor whether her son is well. The doctor saw the son yesterday, when he was fine, but knows that he was killed shortly afterwards. The doctor wants to spare the dying woman the news of her son's death. She utters:

(3) He's fine.

(4) I saw him yesterday, and he was fine.

While (3) outright *says* something false, (4) merely *implicates* something false, while its explicit content is true and, as such, unobjectionable. Many philosophers are interested in the question of how the moral status of misleading compares to that of lying. For our purposes, we can sidestep this question and simply recognize that one of the main purposes of misleading is to 'trick' hearers into believing or acting on certain contents without the speaker being committed to those contents.

Finally, consider insinuation. To illustrate the phenomenon, consider a few well-known examples:

(5) A driver stopped for speeding: "I'm in a bit of a hurry. Is there any way we can settle this right now?" (from Lee and Pinker 2010)

(6) In a letter of recommendation: "Mr. X's command of English is excellent, and his attendance at tutorials has been regular." (from Grice 1975)

(7) A realtor to potential buyers from a different religious background: "Perhaps you would feel more comfortable locating in a more... transitional neighborhood, like Ashwood?" (from Camp 2018)

In (5), the speaker insinuates that they'd like to offer a bribe, in (6), the letter writer suggests that Mr. X is a poor job candidate, and in (7), the realtor communicates that the potential buyers aren't welcome in the neighborhood. What makes these sentences cases of insinuation, according to Elisabeth Camp (2018), is that they communicate "beliefs, requests, and other attitudes 'off-record,' so that the speaker's main communicative point remains unstated" (p. 42). As she correctly points out, however, this analysis wouldn't distinguish insinuation from mere conversational implicature, as in (8):

dogwhistles. Due to space constraints, we won't give the matter the attention it deserves, but we'll assume that deniability is a strategy available to at least all intentional dogwhistles, whether covert or overt. For a detailed discussion, see Saul (2018) and Khoo (2017).

(8) *A* (standing by the side of the road): I'm out of gas.
*B*: There's a service station two blocks up State Street.

What gets at the intuitive difference between cases of insinuation and mere implicature, then? According to Camp (2018),

> [w]hat is distinctive of insinuation is that if $H$, or someone else who overhears the conversation or hears an indirect report of $U$, explicitly attributes [speaker meaning $M$] to $S$, then $S$ is *prepared and able to coherently deny* [$M$]. (emphasis added, Camp 2018, p. 45)

Although Camp is hesitant to give a definition of insinuation, she stresses that the "phenomenon of *implicature with deniability* clearly lies at its core" (Camp 2018, p. 46) and is what makes insinuation practically useful, in that it allows speakers to "make a conversational move without paying the conversational costs" (p. 47). Note that if we accept Camp's characterization of insinuation, all types of strategic speech that bank on deniability through indirectness would be special cases of insinuation—including dogwhistles and misleading speech.

Let us now contrast types of strategic speech with subtweeting. Just like in the case of insinuation, subtweeting doesn't only involve implicature, but intentionally exploits the deniability of contents it alludes to indirectly. Thus, subtweeting is an *instance* of insinuation. What's *special* about subtweeting, though, is that whatever is insinuated *must* pertain to the target's identity. To see this, consider (9):

(9) @Prof_Leslie: "people are asking me how Heidi's talk went....uhhh.... all I can say is her slides were beautiful lmao"

Although (9) involves an implicature that Leslie could possibly deny when pressed,[12] it isn't a subtweet, simply because the target of the subtweet is made explicit. This reconfirms a point we made earlier: while subtweets regularly leave additional parts of their content implicit, a central—if not essential—feature of subtweets is that their *target* isn't made explicit. But this gives rise to an important question: Why does Twitter 'select for' target deniability as one of its distinctive forms of strategic speech? We will discuss this question in §2.3, after we finish comparing subtweeting to others forms of such speech.

Subtweets and dogwhistles also have important parallels. An under-appreciated aspect of subtweets is that, much like dogwhistles, they send messages in a 'multilayered' way. Recall that dogwhistles generally allow for two interpretations: an innocuous interpretation targeted toward a general audience and an objectionable interpretation targeted to an 'in-group.' In an important way, a subtweet's implicit content is intended to be recoverable for, and addressed

12. E.g., "Omg!!! I totally didn't mean to throw shade on the content of the talk! I've constantly been complaining that senior profs make the worst slides. I meant to say that it takes a responsible junior prof to finally get minimally appealing slides!"

toward, a subset of the audience that is already "in the know" and thus capable of recovering the objectionable content. Call this subset of the audience $Aud_{in}$.[13] In §2.3, we will see that, much like in the case of dogwhistles, $Aud_{in}$ helps users navigate multiple audiences on Twitter. In *contrast* to dogwhistles, however, authors of subtweets typically intend for (at least some) members of the out-group to recognize *that* the tweet is a subtweet, even if these members don't (yet) have the resources to recover its hidden content. Call this subset of the audience $Aud_{med}$. Subtweets often do succeed in signaling to a subset of the audience that they're subtweets.[14] If their curiosity is sufficiently piqued, these members of $Aud_{med}$ can initiate the necessary inquiries to find out whom the tweet was supposed to target.

We haven't yet discussed all the ways in which subtweets are like dogwhistles. Like dogwhistles, (at least some) subtweets are also directed at parts of the audience—the 'outgroup'—that can neither recover the sutweet's implicit content, nor recognize that the subtweet is a subtweet. Rather, this group assigns an unobjectionable, innocuous interpretation—usually the tweet's literal content—to the tweet. Call this subset of the audience $Aud_{out}$.[15] Subtweets that are expressed via platitudes or generalisms are especially apt for these purposes:

(10) If you see a colleague in the hallway of your office and don't say hi, you're an anti-social asshole.

(11) Sometimes the people you most admire become the people you least respect.

(12) Everyone you meet might be fighting a battle you know nothing about. Be kind.

Members of $Aud_{out}$ can, for example, simply assign the literal content of (10)–(12) as their respective interpretation. In this case, (10)–(12) might simply be read as making a general pronouncement or expressing general 'wisdom.'[16]

Twitter as a conversational platform is particularly well-suited to give rise to an audience such as $Aud_{out}$ that assigns a literal, innocuous content as the communicative intention behind a subtweet.[17] In order for the literal content to be available as assigned speaker meaning, it must be possible to construct it as a minimally relevant contribution. However, it's very unclear what counts as 'minimally relevant' in the context of Twitter. Moreover, the standards of relevance for initiating a 'conversation' on Twitter—i.e., tweeting something—are fairly low.[18] This, in turn, has important consequences for deniability. As we have seen in the last

13. In the context of subtweets, $Aud_{in}$ doesn't necessarily consist of people who have a favorable attitude towards the content of the subtweet or its author. In some cases, it might just consist of the subtweetee.

14. Indeed, even an algorithm was relatively successful at subtweet detection (Segal-Gould 2018).

15. In cases like (1), the audience might be exhausted by $Aud_{in}$ and $Aud_{med}$, and $Aud_{out}$ is the empty set. But in more mundane cases of interpersonal drama, like (2), $Aud_{in}$ will be considerably smaller, and $Aud_{out}$ considerably larger.

16. We return to complications raised by these cases in §2.4.

17. This is true even for non-platitudinous subtweets.

18. This relates to problems about how social media users construct models of conversational contexts: How do we identify the Question Under Discussion, set of mutually accepted propositions, plans in place, etc., especially

section, when a speaker exploits deniability, they pretend that the intended content of their tweet does not involve target $T$. But it has been pointed out by many theorists (Camp 2018; Lee and Pinker 2010; Goffman 1955) that plausible deniability has its limits: it only succeeds if you can point to a "possible" or "virtual" audience that would reasonably and "sincerely employ alternative assumptions $I'$ to derive $[M']$" (Camp 2018, p. 51). The existence of $Aud_{out}$ makes this an easy task.

$Aud_{out}$ gives rise to an important connection between subtweeting and misleading. Concretely, by being made to believe that the tweet at issue is not a subtweet, and that the innocuous, literal content of the tweet corresponds to its communicative intention, $Aud_{out}$ seems to be actively misled. As a result, members of $Aud_{out}$ might engage with it in a way they might not want to if they were aware of the tweet's function. In §3, we will develop this important insight further, and show how the possibility of misleading can give rise to morally problematic aspects of subtweeting.

In this section, we drew on theoretical insights about different types of strategic speech to illuminate how subtweeting works. At the same time, we highlighted the differences between them, showing that subtweeting is a *distinctive* phenomenon with its own communicative mechanics. Thus, the contrastive exercise in this section showed us a novel way in which deniability can work. In the next section, we will show how subtweeting's communicative features interact with features of Twitter's platform design.

### 2.3 *Subtweeting and Platform Design*

A medium for communication crucially shapes *how* we communicate.[19] This is no less true for Twitter. Drawing on the insights from the previous subsection, here we explore how Twitter's designed speech environment relates to the communicative practice of subtweeting. More concretely, we shed light on the aspects of Twitter's architecture that encourage subtweeting as a strategic practice, discuss how subtweets help navigate multiple audiences, and offer an explanation for why Twitter gave rise to a form of indirect speech that systematically hides target identity. It's no accident that the phenomenon referred to as *subtweeting* arose on Twitter.

A common trope in popular discourse about the internet is that it 'doesn't forget.' This idea seems exaggerated when applied to 'the Internet' as a whole. But it might be a little closer to truth when it comes to Twitter. Twitter is, in many ways, both *permanent* and *public*. As digital researchers Eden Litt and Eszter Hargittai put it, "in comparison to offline interactions, communication tends to be more persistent, searchable, archivable, and shareable" (Litt and Hargittai 2016). While it is possible to delete tweets, the threat of screenshots constantly looms, and some tweets are archived by time machines. By contrast, everyday 'real life' conversations

given that we don't know the 'participants' of our 'conversations'? See Goldberg (2021) for an insightful discussion.

19. For some relevant literature, see Tufte (2003a,b, 1990); Nguyen (2021a); McLuhan (1994).

are infrequently recorded. This permanency of Twitter stands in a symbiotic relationship with its publicity. Since Twitter is a predominantly public platform, most tweets can, in principle, be seen by anyone.[20] This point has also been emphasized by media scholars Alice Marwick and danah boyd:

> When an individual's account is public, *anyone*—with or without a Twitter account—can read their tweets through the site, RSS, or third-party software. [...] Additionally, it is not uncommon for people to forward tweets via email or by copying and pasting them into new communication channels. Furthermore, various tools allow users to repost tweets to Facebook, MySpace, and blogs. (our emphasis, Marwick and boyd 2011, p. 117)

Yet the more people can see your tweets, the more likely it is that your tweets stay permanent—either by increasing the likelihood of direct 'archiving' (e.g., screenshotting), or simply in virtue of entering collective memory. At the same time, the permanency of a tweet makes it easier for it to reach a wide audience through the mechanisms Twitter has designed for this purpose: retweeting and quote tweeting.[21] By contrast, again, everyday conversation are usually not public, and the default on sites like Facebook and Instagram is to have a private profile; typically, only people one accepts as friends or followers can see one's posts.[22] Yet one's potential audience for Twitter can go far beyond one's followers, or even Twitter users (cf. Marwick and boyd 2011).

As we see it, all this has vast consequences for the deliberative statespace of Twitter-users. The threat of unsavory tweets being shared, and of the Tweeter being subsequently shamed, strongly incentivize the creation of messages with ambiguous or vague meanings, allowing objectionable readings to be denied.[23] In addition, due to Twitter's public nature, a Tweeter

20. Only around 13% of Twitter profiles are private as of 2019 (see https://medium.com/pew-research-center-decoded/how-public-and-private-twitter-users-in-the-u-s-d536ce2a41b3).

21. This vast potential for repeated sharing also differentiates Twitter from other potential loci of subtweet-like phenomena—such as rap songs, 'blind items' in gossip columns and podcasts (where clues about the target of gossip are offered but identities obscured), and romans à clef (novels where real people or events appear under fictitious names). Compare, though, the infamous kidney story from the NY Times: https://www.nytimes.com/2021/10/05/magazine/dorland-v-larson.html, where a story essentially about the literary equivalent of subtweeting did become viral.

22. Some analogues of subtweeting can occur on other social media sites—most saliently, Facebook, where it goes by the name 'vaguebooking.' Vaguebooking is defined as "the practice of making a post on social media, primarily Facebook, that is intentionally vague but highly personal and emotional" (Dictionary.com). Examples include users posting statuses like, "I can't even right now," and "Why can't I catch a break?" (For more examples and discussion, see Parkinson (2014) and Nester (2015).) Unlike subtweeting, vaguebooking is almost univocally seen as desperate and attention-seeking, and it does not necessarily have a specific target. As Elle Hunt, writing for *The Guardian* notes, "A subtweet is far more pointed, and about or in response to a specific subject. Exactly what or who is not explicitly stated but it's often not hard to figure out" (Hunt 2017). We can also imagine analogues of subtweets occurring on Instagram or Tik Tok, yet these platforms are not primarily text-based, and hence it would be more rare.

23. Interestingly, in certain contexts, there might also be pressures *against* ambiguity and vagueness when the likelihood of deliberative misinterpretation is high. However, this kind of pressure might be more relevant when

must anticipate a variety of audiences, who bring to bear different beliefs, assumptions, conceptual schemas, and values. While in-person contexts also can require communicating to distinct audiences simultaneously, this need is intensely magnified in online settings. This, in turn, has important consequences for our epistemic situation. In some online settings—like Facebook, where your audience is usually your curated list of friends—you can reasonably anticipate what some of the beliefs of your audience members are, and craft your posts accordingly. In other online contexts—like Twitter—you can be completely in the dark about the beliefs and values of your audience. First, the sheer number of expected readers can be far greater. Second, the potential consumers include potential *future* readers, who may read the text in a different context or with different social norms in the background. While we can usually make good guesses about the values and background beliefs of our audience in in-person contexts, Twitter makes our epistemic situation as discourse participants much more dire. Indirectness can be a solution to the potential costs of this epistemic 'veil.' Keeping controversial contents vague or inexplicit, and therefore deniable, can be a welcome preemptive measure against backlash from audiences whose reactions to your tweets you can't anticipate.

Within media studies, this well-known problem has been discussed under the umbrella notion of the *imagined audience* (Litt 2012; Litt and Hargittai 2016; Marwick and boyd 2011). The imagined audience serves as a guide for determining what is appropriate to communicate with one's audience.[24] But, as Marwick and boyd observe, "it is virtually impossible for Twitter users to account for their potential audience, let alone actual readers" (Marwick and boyd 2011, p. 117). Hence, to avoid *context collapse*—or the flattening of multiple audiences into one—users must "negotiate multiple, overlapping audiences by strategically concealing information, targeting tweets to different audiences and attempting to portray both an authentic self and an interesting personality" (Marwick and boyd 2011, p. 122).[25] As we saw in §2.1, subtweets are an excellent means to navigate communicating with distinct audiences at once while avoiding context collapse, as they partition their audience into three subsets: $Aud_{in}$, $Aud_{med}$, and $Aud_{out}$. By coding their messages via subtweets, users can converse with those 'in the know' ($Aud_{in}$), even cultivating a sense of digital intimacy (Thompson 2008) by alluding to their private lives, while outsiders ($Aud_{med}$) can merely speculate about—or attempt to piece together—the subject matter. All the while, the existence of $Aud_{out}$ increases the possibilities for plausible deniability in the case of pushback.

A closely related aspect of Twitter highlighted by media scholars is the need to create bits

the intended speaker meaning is *not* risky, or less risky than strategic misinterpretations. In these cases, avoiding vagueness mitigates risk. In the case of subtweets, the intended meaning is risky; hence, indirectness mitigates risk. Thanks to Sandy Goldberg for bringing this to our attention.

24. Hence, the imagined audience plays a similar role as the representation of the participants in a conversation under a Stalnaker-Roberts model of conversation.

25. See also Lucy McDonald, "Context Collapse Online" and Regina Rini, "Context Collapse and Pop-Up Communities: How Social Media Makes Its Own Norms" in this volume and Frost-Arnold (2021) for further, nuanced discussion of context collapse. See Nguyen (2021b) for a public philosophy piece on how context collapse on Twitter impacts intimacy.

of *privacy* in a platform that's public per design (Marwick and boyd 2014).[26] Insofar as there is a genuine need for private spheres in online discourse, speakers will use different strategies to satisfy this need in environments that don't afford it. Applied to Twitter, one way to achieve a higher degree of privacy is to create a private 'alt-account' that only allows for a curated set of followers, or to create a 'Twitter Circle,' a feature that enables that a subset of your tweets is only visible to an approved list of followers. But another, perhaps more efficient, way to achieve (semi)-privacy is subtweeting. Because subtweeting partitions the audience in the way outlined above, only the audience with the relevant background knowledge, $Aud_{in}$, will be able to seamlessly recover the intended content of your subtweet. Since the subtweeter will be able to anticipate who has this background knowledge, they will be better at predicting the in-group than the entire audience.

The primarily public nature of Twitter also helps explain why the platform gave rise to a systematic subcategory of insinuation that conceals *the target* in particular, i.e. subtweeting. If you want to get across risky contents about another individual in real life, there's an easy way of doing this: by talking about them behind their backs. But because the dominant form of interaction on Twitter is public, Twitter users are deprived of this possibility. So instead of concealing *the conversation* from the target as in instances of gossip, subtweeting attempts at a similar effect by concealing someone's *identity*, but not the conversation. In short, because Twitter removes the privacy we need to gossip, Twitter users resort to other ways to criticize or mock someone.[27] But this can't be the whole story. After all, if it was merely for lack of privacy, Twitter users could simply resort to direct messages in order to socially reprimand their targets. When we subtweet, we *want* to criticize or mock someone publicly. We *want* people—optimally, many people—to know that an individual has done something mockable or criticizable, receive validation for it, and perhaps even show the target that they are being disapproved of or ridiculed. While Twitter takes away privacy from users, it *gives* them publicity. This publicity is exactly what is needed for the desired public criticism. A subtweet can give you the desired publicity, while simultaneously decreasing the likelihood that you're perceived as moral transgressor, disproportionate punisher, or otherwise inappropriate (see §3), and increasing the likelihood of Twitter-related rewards (e.g., likes and approving comments). Thus, with subtweeting, we can reap many of the benefits of public criticism *and then some*, without the associated costs.[28]

This brings us to the next point. Other features of the Twitter platform and its incentive structure makes subtweeting both prevalent and predictable as type of speech that would emerge. Twitter is explicitly designed to be conversational. It allows for, indeed encourages,

26. Thanks to Josh Habgood-Coote for comments related to this point.

27. This explains why subtweeting is sometimes seen as the "internet equivalent of talking about someone behind their back" (Parkinson 2014).

28. Interestingly, this is why analogues of subtweeting all happen on public media: newspaper, roman à clef, Facebook, and so on. However, subtweeting has strategic aims that aren't present elsewhere (e.g., *both* publicity and instant validation), and this explains its ubiquity on Twitter.

real time feedback and conversation. But, as Thi Nguyen emphasizes, this results in Twitter *gamifying* communication "by offering immediate, vivid, and quantified evaluations of one's conversational success" (Nguyen 2021a, p. 411). Tweets are often crafted so as to provoke response and approval, for tweets are 'scored' by their number of retweets and likes. As one *Slate* article emphasizes, "the more drama you can bring to yourself, the more followers and favorites you acquire."[29] Subtweets pique curiosity and function similarly to gossip. Eliciting curiosity and conveying gossip is a reliable way to enhance engagement on social media, hence scoring Twitter 'points.'[30] Hence, insofar as subtweets frequently provoke curiosity about their target—as they surely do and, indeed, via the creation of $Aud_{med}$ are often intended to do—we should expect subtweets to facilitate engagement.[31] Combined with what Nguyen (2021b) calls the "gratifications of shared outrage," subtweets can provide an effective way for enhancing a user's Twitter 'score.'

We've highlighted certain design elements of Twitter that, as we see it, encourage the emergence of subtweeting as a systematic discursive practice. This doesn't mean, of course, that all risky content will be expressed via subtweets or other forms of indirect speech. In fact, you won't need much time to find a tweet that's extremely offensive, foolish, or otherwise risky. When and whether we subtweet will depend on a host of contingent factors that enter your practical deliberation. These include, but aren't limited to, the name under which you tweet, your individual degree of risk-aversion, independent principles regarding concealed speech, how sympathetic one's audience might be, and expected benefits of explicit speech. These factors, however, are too messy for us to say anything general about. For now, we leave it at one generalization: *If* you want to get across a risky content about an individual while avoiding liability, subtweeting gives you *a* strategy to achieve this.

To sum up, it is unsurprising that subtweeting emerges on Twitter: subtweeting gives users a way to publicly criticize or mock someone while minimizing associated risks by concealing the tweet's target. These risks are particularly high in public online platforms, such as Twitter. In addition, the platform itself incentivizes subtweeting as a means of navigating multiple audiences and enhancing engagement. These features make the act of subtweeting particularly attractive to participants in a Twitter-game. It is no surprise that sub*tweeting* emerged.

29. https://slate.com/technology/2016/07/subtweeting-looks-terrible-on-you-yes-you.html.

30. For relevant empirical research, see for instance Alicart et al. (2020) on how gossip information increases reward-related neural activity and Sharron and Abraham (2015) on how curiosity plays a key role in motivating individuals to engage with an object or site. While this is more speculative, various studies emphasize the role of high-arousal emotions in increasing virality (Berger and Milkman 2012; Brady et al. 2017). Insofar as curiosity is also a high-arousal emotion in the relevant sense (Berlyne 1954), we have further reasons to expect subtweets to elicit engagement.

31. This is true even if subtweeters are sometimes evaluated negatively, in part for communicating indirectly (Edwards and Harris 2016). However, people may under-report how much they enjoy subtweets, insofar as mocking or being mean to others is taboo (cf. Peng et al. (2015)). Judging negatively is not incompatible with engaging positively. Moreover, it's worth noting that the Edwards and Harris study only used one example of a negative subtweet in generating their findings: "Thanks to a certain person for backstabbing and completely ruining my day. People like that are pathetic." This is not particularly clever or subtle.

### 2.4 *Complications: Meanings, Contexts, Intentions*

Complications abound. Is any tweet that indirectly alludes to someone in a critical way a subtweet?[32] Can something appear to be a subtweet when it in fact is not? How precisely do the user's intentions fit in when determining whether something counts as a *subtweet*? These questions raise complications in giving an account of subtweets in general. We'll foreground them by discussing two types of difficult cases.

First, what should we say about tweets that are intended to be about something *general*, yet appear to onlookers as being about someone specific?[33] Consider, for example:

(13) I hate it when people don't show up to events on time.

(14) "People who write in English but include untranslated passages from German/Latin/Greek/French/etc. are dicks."[34]

Suppose further that the authors of the tweets don't have anyone particular in mind they intend to target. Do these count as subtweets? There is good reason to think not. After all, we want to be able to express attitudes and sentiments about a *type* of behavior or person. In doing so, we may be seeking support, attempting to discourage the behavior, or simply venting.

Even stronger, we might think: even if a particular person *prompted* a tweet, that doesn't necessarily mean it's *about* them.[35] Instead, the user may want to make a general point. Their identity might not be relevant to what the user is attempting to get across, or the user may want to retain anonymity for other reasons. This is especially obvious when the user doesn't even *know* the identity of the person who prompted the tweet. Imagine, for instance, that a random man catcalls a woman, and she writes:

(15) In shocking news, harassing women isn't a good way to flirt.

It would be far too strong to claim that she is subtweeting him. Similarly, when Lil Baby criticized men who subtweet other men, he was able to avoid the charge of hypocrisy simply by adding:

(16) "not a subtweet I'm being direct to all men."[36]

32. The jury is out on whether subtweets can be positive. Edwards and Harris (2016) assume they can, but it's controversial. At the very least, positive subtweets are rare. See this informal Twitter poll by Jonathan Ichikawa for some disagreement, with the majority of respondents voting that they must be negative—or, at least, cannot be overly positive. Similar questions arise for gossip. Note that it's possible that a content is risky even when positive.

33. Thanks to Patrick Connolly for raising this question.

34. Tweet by Michael Hannon, included with his enthusiastic blessing: https://twitter.com/m_j_hannon/status/1569683673942736899.

35. What it means to be *about* someone is a difficult question. See Yablo (2014) for a book length treatment.

36. https://twitter.com/lilbaby4PF/status/1196251324008062976?ref_src=twsrc%5Etfw. Original tweet here: https://twitter.com/lilbaby4pf/status/1196183580139016192.

His denial that it was a subtweet is, at the very least, coherent and not automatically self-undermining.

Second, what should we say about cases where a subject is not mentioned simply because their identity is already part of the discourse? In such cases, the user mocks another user without mentioning them *not* because they want to exploit deniability, but because their intended audience already knows whom they are talking about. (If they don't, perhaps the tweet is not for them; as we saw, omitting details can be a way to avoid context collapse.) This frequently happens with tweets about the 'main character on Twitter that day' (i.e., Tweets about a person that has received a lot of attention on Twitter or one of its sub-communities on a given day), or tweets that take on a meme-like character, referencing an original tweet in order to mock it.[37] Call these 'main character' and 'meming' tweets, respectively. Consider, for example, various tweets obliquely referencing Nigel Warburton's

(17) "If you can't say it clearly, quit philosophy."[38]

Reactions included,

(18) "If you can't solve the problem of induction, you should quit philosophy,"[39]

(19) "If you can't typeset it beautifully, quit philosophy."[40]

Do such cases nonetheless count as subtweets?

We think all of these cases foreground the role of intentions when delineating subtweets. Let's start with tweets that are intended to be about something general but look like subtweets. Subtweeting is a pragmatic phenomenon; we use the evidence provided by the speaker to recover speaker meaning $M$. But according to standard Gricean accounts of speaker meaning, communicative intentions are exactly what determine speaker meaning. More precisely, for a speaker to mean $M$, she must intend for the hearer to come to believe $M$ *and* for the hearer to recognize this very intention. But in cases like (13)–(16), the speaker neither has the intention to talk about a specific person (and hence to subtweet), nor the intention for the hearer to recognize this intention. In other words, the content of $S$'s communicative intention $M$ behind the tweet does not involve target $T$. By hypothesis, the authors of (13)–(16) only *intend* to make a general point. Hence, $T$ isn't part of speaker meaning $M$, and (13)–(16) are, if anything, only pseudo-subtweets.[41]

Nevertheless, it is an interesting question whether the authors of (13)–(16) can still be held responsible if their tweets are understood as subtweets targeting individuals.[42] As has often

37. Thanks to Regina Rini for prompting us to think more about 'main character' tweets.
38. https://twitter.com/philosophybites/status/1375377478332604418?s=07. Included with Warburton's permission.
39. https://twitter.com/DailyNousEditor/status/1375863509377028096?s=07.
40. This example is from a tweet that was since deleted.
41. Note that some cases of generic tweets can both intend to make a general point and have a particular target in mind, and hence can still be subtweets.

been pointed out, it is the speaker's responsibility to make available enough information for their desired interpretation. This is demonstrably not the case in subtweets that are *systematically* misread as subtweets, whatever the author's intention may be. In these cases, due to communicative negligence, the authors of the tweets might be *taken* to have subtweeted, and both they and the suspected target might be treated accordingly.[43] Note, also, that the existence of generic pseudo-subtweets offers a further route to deniability to actual subtweeters. Even when a tweet *is* about a specific person, the user can deny this, instead insisting that they were talking about a generic phenomenon despite having a clear target in mind. Subtweeters can just exploit the existence of pseudo-subtweets to achieve their aims.[44]

This addresses the first complicating case. The second case is more tricky, and we think intuitions will come apart as to whether or not these cases are subtweets. We won't take an ultimate stance as to whether or not meming tweets or 'character of the day' tweets are subtweets. However, we think there is an explanation for what drives either judgment, which itself contains lessons about subtweeting's nuances.

Let's start with the intuition that 'meming' and 'main character' tweets like (17)–(19) are *not* subtweets. What's behind this intuition? First, it seems that in many of those cases, the speaker is not typically—or primarily—motivated by a desire to exploit deniability. As we've seen in §2.1, this desire seems to be crucial to distinguish between forms of insinuation (including subtweeting) and mere implicature. Rather, the reason that the mocked target isn't named seems importantly different. In the case of memes, for example, making the contextual information required to understand the meme explicit would ruin precisely what's funny about memes and the shared intimacy created through them (Nguyen 2021b; Cohen 1999). Moreover, as we will point out in §4, meming tweets can work to resist or subvert problematic norms. Thus, the main strategic function behind subtweets and many meming tweets seems to differ: the former, though not the latter, aim to exploit deniability. This probably drives the intuition that meming tweets are not subtweets. Similar points can be made for 'main character' tweets.

However, 'meming' and 'main character' tweets also share many commonalities with subtweets. Here are some obvious ones: there is an intended specific target, and the tweet is, at least often, about the target in a way that's mocking or critical. In addition, they often seem motivated in part by a desire to avoid context collapse (§2.3). Because of these important commonalities, it makes sense to think that these sorts of tweets are subtweets, too. But there might be another, more subtle commonality. Risk-reduction is one of the key aims

42. Indeed, this question seems to be an instance of the more general question, namely: what determines the speaker's responsibility in cases in which the speaker and audience bring different normative expectations to bear on the speaker's act? Thanks to Sandy Goldberg for this point.

43. See Saul (2002) for interesting discussion.

44. Relatedly, note that the ambiguity between subtweets and (for example) generalities is a corollary of the fact that subtweets exploit deniability; this deniability is only possible insofar as the same tweet can be given a different, unobjectionable interpretation. It follows that there *can* be tweets that look like subtweets but aren't. The same point applies to other types of strategic speech that bank on deniability.

of subtweeters. So far, we focused on deniability as the mechanism through which subtweets mitigate risk. But subtweets might mitigate risks through other means. For example, not naming someone might make it less likely that the target's followers attack you, or that the target sees your criticism. Avoiding direct confrontation itself can be an instance of risk mitigation.[45] So even if someone who mocks someone else through a meme wouldn't *deny* that their meme is about them, they could still be moved to meme because it allows them to make fun of the target with fewer risks. Insofar as it's plausible to think that users still, at least partially, use meming tweets for this reason, this, together with the other commonalities, might explain the intuition that meming and main character tweets are subtweets.

If we have diagnosed both sets of intuitions accurately, we can end this section with this conjecture: the more likely we think it is that a Tweeter obscures the identity of a tweet's target *because* this mitigates risk, the more we will judge the tweet in question to be a subtweet. Hence, tweets that mock another individual are more likely to be counted as subtweets. By contrast, in some examples of main character and meming tweets, the speaker really just wants to signal that they are up-to-date with the Twitter discourse, not to mock anyone.[46] When this is clear, we suspect that the tweet is less likely to be read as a subtweet.

With an account of the central phenomenon fully on the table, we can now turn to ways in which subtweets can be strategically deployed in (sometimes) problematic ways.

## 3 Deniability and its Discontents

Sometimes, strategic speech and deniability can be deployed in useful and positive ways. We can use it to flirt, to nudge, and to tactfully criticize. This potential for positive usage is no less true for subtweets. As we saw, it can be useful for avoiding context collapse and, as we'll argue below, to resist problematic norms. However, as with strategic speech generally, it can sometimes be deployed for problematic or harmful ends. Here we articulate four ways in which the usage of subtweets can give rise to ethical problems. These involve normative cover, self-defense, harms, and complicity. Then, in §4, we discuss positive uses of subtweets and how their problematic potential can sometimes be outweighed.

**Normative Cover** First, the deniability and inspecificity of subtweets give their authors a publicly acceptable way to perform actions that would otherwise be clearly unacceptable. In other words, it grants them *normative cover* for problematic acts. The communicated content of many subtweets would, if stated explicitly, often constitute a clear norm violation within a given community.

45. Thanks to Neri Marsili for pointing this out.

46. Alternatively, they may just find it amusing and want to contribute their own version. Thanks to Patrick Connolly for this point.

To illustrate, consider again the case of Heidi. Many consider it unacceptable to publicly and unconstructively discredit the philosophical competencies of junior members of the profession. But by covering up their disparaging comment in a subtweet, the author of the subtweet gets a pass for what we would usually condemn. At the same time, the deniability of subtweets prevents them from being adequately contested and the subtweeter from being held publicly accountable. Thus, as with dogwhistles and insinuations, subtweets give their authors a way to make a "conversational move without paying the conversational cost" (Camp 2018, p. 47) and to avoid liability for their norm violation.

Moreover, the subtweet even allows Prof Leslie to take the moral high ground. She can construe her omission of Heidi's name as an act of benevolence: "at least I didn't name her."[47] This defense misdirects attention away from her actual norm violation to how she could've done *worse*, giving her further normative cover—from others, but also from herself. In this respect, it functions similarly to moral self-licensing. In standard cases of self-licensing, an agent appeals to her past good behavior or moral character to justify subsequent immoral behavior: you pat yourself on the back for stealing only $100 from your friend, since you could have stolen $200. Here, something similar occurs: the speaker appeals to the fact that they could've done something worse to justify and neglect criticism from their present behavior.

It might be objected that this feature of subtweets isn't always problematic. To the contrary: the only feasible way for me to, say, complain about my abusive and megalomaniacal boss might by subtweeting them. The conversational costs of talking about my boss directly might just be too high in this case, and subtweeting provides me with some protection from retaliation by those in power.[48] We agree that this instance is an unproblematic case of subtweeting, but we locate the source of this judgment in the *norm* that is operative in the case, not in the act of subtweeting itself. The communicative norm that determines the high conversational costs—*don't criticize your superior*—is independently problematic, as is the high cost that's associated with its violation, which is why we find the 'normative cover' provided by subtweeting unobjectionable in this case. The norm simply protects us from sanction for violating a norm that we already find objectionable.[49] This case is markedly different from Prof Leslie's fictional subtweet. We usually abide by professional norms such as, *Don't put junior colleagues down publicly and unconstructively!* Leslie's subtweet is problematic because it violates a norm we accept, *and* because it does so 'for free'—without accountability or liability. Our point is that *when* there is an objectionable norm violation, subtweeting can prevent us from identifying it.

47. We borrow this way of putting the point from Justin Weinberg. See here: https://twitter.com/DailyNousEditor/status/1590426305941864448.

48. Including, but not limited to, unfair attacks from the target's followers. Thanks to Nikki Ernst for this point.

49. Note that this point generalizes to all types of strategic speech. For example, instances of dogwhistles, code words, insinuation, and misleading can be extremely useful tools in contexts of oppression. Imagine, for instance, Nazi resistance groups using code words to avoid being persecuted.

**Defending Oneself** Second, targets of subtweets can be deprived of the ability to defend themselves.[50] That is, subtweeting is an unfair game in that one party has a power to attack the other without risking being attacked, and the one who is subtweeted is defenseless. As Ian Bogost aptly puts it, "A true subtweet eludes response, because it is so ambiguous as to make response impossible" (Bogost 2015). If everyone has a defeasible right to defend themselves from public attacks, there is a sense in which subtweeting is necessarily morally problematic. To defend oneself, or call out the tweet, would require first outing oneself as the target and centering oneself in the message. However, this risks appearing paranoid at best and narcissistic at worst. ("You're so vain: of course you'd think this subtweet was about you.")[51] Similar worries about centering the target hold for others who may want to defend the target. All of this is exacerbated by the fact that in many cases of subtweeting, the target does not know *for sure* that they are the target. This is true even if they have strong evidence for this given background facts and the content of the tweet.

Protesting can also make you appear overly sensitive to criticism or otherwise humorless; after all, subtweets are often crafted to have the texture of an inside joke. Indeed, this is why even tweets that mention the person by name but without 'tagging' them—what Bogost calls a 'supertweet'—can be difficult to respond to. As Bogost elaborates:

> [Supertweets] prevent the indicted individual from being able to reply without compromising themselves, and in so doing they carve out new room for the supertweeter in a landscape purportedly overgrown with the supertweetee's largess. Just like the political opponent can't reply to the accusation, "I do not know if my opponent in this race is a crook" without repeating (and thereby affirming or at least directing attention to the fact that he or she might be a crook), so the supertweetee cannot address the accusation or indictment without proving [...] that he or she is unwilling or unable to allow other voices to weigh in on the matter without attempting to re-take control of the conversation.

Similar points hold for subtweets, where the target is not even mentioned by name. To attempt to respond to the tweet or defend oneself is to risk appearing self-centered, self-serious, or overly defensive, if not worse. Assuming that there is some defeasible right to defend oneself against public verbal attacks just as much as physical ones, subtweeting will render it very difficult, practically speaking, to exercise this right without great personal costs.

Subtweets disarm their targets in a further way. Outing yourself will let everyone know—including members of $Aud_{med}$ and even $Aud_{out}$—that you are the target of the criticism or

50. Thanks to Junhyo Lee for this suggestion.

51. As Bogost (2015) continues: "Only a neurotic or a narcissist or a paranoiac would ask after 'I see it's jerk day at The Atlantic,' wondering 'if I'm the jerk youre referring to?' or 'I'm not sure if youre talking about my article but if so feedback is appreciated!' Admittedly, there are plenty of those sorts online, but the subtweet acts as cover against any such responses."

mockery. The target might not want to take that risk, especially if the tweet describes them in ways that are embarrassing or mean. Even if the descriptions are false, exaggerated, or cruel, it may take a great deal of effort to correct them once they have become 'attached' to you. In addition, the act of self-identification can itself lend credence to the content of the subtweet.

This doesn't mean that defense is never possible in the face of subtweets—in fact, we've seen targets successfully push back against attacks harbored against them via subtweets.[52] We won't give a full analysis of the conditions that will make such a defense easier, but we suspect that the relative sizes of the audiences will be an important factor. The smaller the size of $Aud_{out}$ and the larger the size of $Aud_{in}$, the easier it will be for the target to defend themselves. Relatedly, we suspect that subtweets are most 'effective'—and can do the most damage—when they are deniable enough to provide normative cover and incentivize not contesting them, but not so vague that they leave too few people in $Aud_{in}$.[53]

**Harms and Hurts** So far, we have highlighted the ways in which subtweets create norm-violations and harms that are not easily acknowledged or redressed. We now want to say more about these harms themselves. Joel Feinberg famously defined a harm as a setback to one's interest, where an interest is something one has a stake in (Feinberg 1987). Clearly, Anatasija has a stake in not being disparaged online (especially by a more senior member of the profession) and this interest is undermined by Leslie's tweet. This is especially clear if Heidi sees the tweet and recognizes herself as the most plausible target, and knows others—namely $Aud_{in}$ and $Aud_{med}$—will too. At the same time, we just saw how Heidi and others will have difficulty defending her and calling out the norm-violation. Subtweets thus can create harms that prevent their own redress.

Interestingly, in some cases, harms associated with subtweets can accrue to someone who's not the *actual* target of a subtweet. This occurs when a subtweet is systematically understood as targeting a different person, $T'$.[54] It is also possible for *non*-subtweets to have effects similar to subtweets. When mere pseudo-subtweets of the kind in (13)–(15) are crafted carelessly and are systematically perceived as actual subtweets, harms may still accrue to the *perceived* target.

Importantly, even when a subtweet does not harm a target, it can still *hurt* them. A hurt differs from a harm in that it does not necessarily involve a setback to one's stable interests. For example, Anastija might feel isolated, upset, anxious, alienated, and embittered. She will likely feel bullied but powerless to do anything about it. Indeed, subtweets often make even

52. They might even do this by firing an amusing and witty subtweet in response, potentially exemplifying an instance of what we'll later describe as subversive 'resistance subtweet' (see §4). For example, Lady Gaga was effectively able to defend herself against Adam Levine. See https://twitter.com/ladygaga/status/381314428861288448?lang=en and https://twitter.com/adamlevine/status/381195819963006976?lang=en, respectively.

53. Indeed, if the target is too clear, then the speaker could perhaps be liable for slander or libel via the doctrine of innuendo. (See https://defamation.laws.com/innuendo.) Thanks to Benj Hellie for bringing this doctrine to our attention.

54. Thanks to Sandy Goldberg for this point.

potential targets feel paranoid, eliciting responses of the form, "Is this about me?" Even when these do not rise to the level of harms in virtue of not setting back a stable interest, they are still important. Causing hurt unjustifiably or gratuitously is typically still ethically significant and criticizable.[55]

There are other hurts that arise from subtweets. For example, the target may feel betrayed by the original poster, but she may also feel betrayed by audience members who engage with such subtweets positively. We turn to their role next.

**Complicity** Subtweets invite public engagement. This can result in the target feeling piled onto, bullied, and, as we saw above, defenseless. Thus, the participating public—due to the existence of $Aud_{out}$, often unknowingly—will become *complicit* in the subtweet's harms and norm-violations. That is, they may *facilitate*, *enable*, or *condone* the harm by engaging positively with subtweets (Mellema 2008). (Perhaps some members of the in-group can also be complicit in failing to denounce the norm-violations, insofar as they are in a position to recognize it.) This will result in them sharing responsibility for perpetuating harms associated with subtweeting. In addition, positive engagement can make it even more difficult to identify the norm violation in virtue of distracting from the implicit content of the tweet.

It is important to point out that, through the complicity they give rise to, subtweets not only harm their targets but also deceive parts of the audience. As we pointed out in §2.2, subtweets actively mislead members of $Aud_{out}$ by 'tricking' them into engaging positively with a tweet in a way they might not be willing to if they knew about the tweet's actual function. Importantly, this engagement can play an additional strategic role: insofar as the subtweet is a reflection of an interpersonal conflict, the positive engagement can give the author of the subtweet a way to gain leverage in this conflict. For example, positive comments on a subtweet can signal to the subtweet's target that others are in agreement with whatever criticism is behind the subtweet. This can also trick the target into thinking that the $Aud_{in}$ is larger than it is, thus exacerbating their perceived harm and suffering.[56] Thus, not only are members of $Aud_{out}$ misled, but instrumentalized.

It is instructive to see that an analogous point holds for dogwhistles. Imagine that parts of an audience don't know that 'inner city crime' is a dogwhistle for 'black crime.' They thus enthusiastically cheer on a politician without fully understanding what they are thereby supporting. The analogy between dogwhistles and subtweeting is unsurprising, because dogwhistling partitions the audience in a similar way. But the analogy contains an important lesson: partitioning of different audiences, in both subtweeting and dogwhistling, plays important strategic roles that go beyond mere risk mitigation. Moreover, observe that deniability plays a crucial role in complicity of the *knowing* audience in both cases: members of $Aud_{in}$ (and $Aud_{med}$) may themselves want to exploit the normative cover conferred by deniability.

55. Thanks to Christa Peterson for noting the importance of talking about hurtfulness, not just harms.
56. Thanks to Sandy Goldberg for this point.

When pressed, they might claim they did not understand what they were committing to when in fact they did. Deniability can thus provide normative cover not only to the speaker but also the audience, allowing them to avoid the appearance of complicity or moral infraction.

## 4 The Value of Subtweets

In the last section, we pointed to several ethical problems that subtweets can generate. However, the point of this paper is not to advocate for a blanket prohibition on subtweeting. This would be not only too quick but also unwarranted, despite the fact that some writers have suggested that people should stop subtweeting altogether.[57] But we think there are many contexts in which subtweeting is not only justified, but can even be used for the good. In this section, we point to some of these contexts.

Some of these contexts were already alluded to earlier in the paper. For example, in §2.3, we have seen that subtweets can offer a means to avoid context collapse and thereby create some sense of privacy. Thus, subtweeting can serve as a strategy to get rid of some of the constraints imposed by Twitter as a communicative platform. This also applies when subtweeting is used to address concerns surrounding data privacy more generally, including to thwart algorithms (Tufekci 2014). For example, the MRS Delphi Group's report on privacy notes that teens use "sophisticated means to manage privacy, relying on social coding... This is often in the form of 'in-jokes'" (Strong 2015, p. 37). This 'dirties' the data. While the report focuses on 'vaguebooking' (posting vague statuses on Facebook), we take it that it could easily take the form of subtweeting as well. Of course, there might be other ways of avoiding exploitation by algorithms, such as screenshotting, not tagging, or altering the spelling of relevant information so that it is not easily searchable. However, subtweets provide far more deniability than these alternatives, which can be particularly important in political contexts such as protests. For example, subtweets were routinely used during the Gezi Park protests in Turkey (Tufekci 2014). Relatedly, subtweeting can be a way to avoid giving the target more attention and name-recognition (Tufekci 2014) or to avoid amplifying problematic speech.

Our example of subtweeting the megalomaniac boss in §3 foregrounded another value of subtweeting. Sometimes, the only way to criticize or vent about some behavior is by subtweeting the target. In these cases, subtweeting *prevents* harms, because it protects from sanctions for violations of *problematic* norms. Thus, cases like this show that sometimes, there are legitimate reasons to minimize communicative risk—for example, when it creates benefits that

57. For example, *The Washington Post*, reporting on research by Edwards and Harris (2016), claims that "you should pretty much never subtweet unless you really, truly don't care about what people think of you." (https://www.washingtonpost.com/news/the-intersect/wp/2016/06/06/study-confirms-what-you-always-knew-people-who-subtweet-are-terrible/.) The *New York Times* even issued a company-wide memo announcing a Twitter 'reset,' stating that "tweets or subtweets that attack, criticize or undermine the work of your colleagues are not allowed." (https://www.mediaite.com/news/nyt-declares-twitter-reset-in-leaked-memo-warning-of-damage-to-journalistic-reputation-subtweets-are-not-allowed/).

clearly outweigh its potential harms.

Interestingly, some subtweets go beyond a mere protective function. Sometimes, subtweets can even be deployed to actively *resist* and *subvert* problematic norms. Consider a fictional tweet stating:

(20) If you can't publish five papers a year, you're not working hard enough.

Arguably, this tweet conveys problematic norms regarding overwork and hyper-productivity. (Not to mention, it is false, given the role of many other factors.) Suppose further that the person who posted (20) is a very prestigious professor at a top research university. Now, imagine that other Twitter users immediately begin to mock this tweet through riffs, essentially turning it into a meme. They might write things like:

(21) If you're sleeping more than 2 minutes a day, you're not working hard enough.

(22) If your eyes aren't bleeding from staring at your screen for 23 hours, you're not working hard enough.

These tweets bring out the absurdity of the initial tweet. Insofar as we think they are subtweets (see §2.4 for discussion),[58] it is easy to see how they can play a core role in joint acts of subversion and resistance.[59]

Hence we we might want to distinguish between two different types of subtweets: call them *ad hominem* and *resistance* subtweets.[60] Ad hominem subtweets are most of the tweets we discussed in this paper (e.g., our fictional tweet by @Prof_Leslie in (2)), whereas resistance subtweets include the above examples (21) and (22), as well as (18)–(19). What's distinctive about resistance subtweets is that they aim at subverting problematic norms and behaviors; often those that have been implicitly promulgated by a prior tweet or more general discourse. Tweets with a meme-like character, like the ones we discussed in §2.4, seem particularly well positioned to accomplish this. A resistance subtweet can help undermine problematic norms in four ways. First, it underscores how absurd the original tweet was, effectively *reductio*-ing it. Second, it undermines the tweeter's standing by parodying and ridiculing him. Third, by not mentioning someone's name, the target does not receive attention or engagement they might find desirable. Fourth, due to their humorous or meme-like character, they are arguably more effective at subverting a problematic tweet or norm than offering a detailed, erudite list of the ways in which the tweet is problematic. This is especially true given the aforementioned role of humor in cultivating intimacy with those who laugh with us.

58. Although the conclusions we can draw from this are quite limited, this extremely informal Twitter poll indicates general agreement about this, with some detractors: https://twitter.com/eleonoreneufeld/status/1591151271758925824?s=46&t=MNvUMjalCAbn3G0PORuokA.

59. On subverting uses of subtweets, see Jessica Pepp, Eliot Michaelson, & Rachel Sterken, "On Amplification" in this volume.

60. Thanks to Renée Jorgensen for this point and for the suggested distinction and terminology.

We want to end this section by noting three things. First, our view correctly predicts that the *more* specific and the *less* deniable a subtweet is, some of its potential for harm is mitigated. For example, when subtweets are uniquely identifiable and the target is shared knowledge, then the target has more resources to defend herself and the normative violation—should it exist—is apparent. Moreover, there is less risk of complicity. This is not to say that such subtweets are all things considered better; for example, they still constitute bullying, and it can be overall worse for the target's identity to be shared knowledge. Our point is merely that more inspecific and less deniable tweets create *distinctive* harms, even if they are not overall worse. If the alternative to subtweeting is aggressive tweeting, then subtweeting of course may still be better. Second, even when there are benefits to subtweeting, these often must be weighed against the potential harms and ethical problems we've discussed. Third, although we grant that resistance subtweets may be unproblematic, it can be difficult in practice to distinguish resistance subtweets from *ad hominem* ones, which have been our primary focus here. A single subtweet could be both. Hence, we shouldn't be too quick to let apparent resistance subtweets off the hook.

## 5 Practical Upshots and Future Work

This chapter developed an account of subtweeting. We examined the complex communicative profile of subtweeting and the mechanics through which it achieves its strategic aims. We also looked at some of the moral dimensions of subtweets and showed that, while their uses are often problematic, they can also be deployed in valuable ways. While we hope that our chapter has contributed to a richer understanding of a communicative phenomenon that hasn't received much attention, we think it also holds a deeper practical lesson. In order to illustrate it, it will be useful to look at discussions surrounding another social phenomenon: gossip.

Gossip does not have the best reputation. At the same time, while the practice of gossiping has long been theoretically under-explored, recent theoretical discussions have shed light on important social roles for gossiping (Adkins 2017; Alfano and Robinson 2017; van Niekerk 2008; Rose 1984). In fact, gossip is a highly norm-governed activity, and most of us have acquired implicit mastery of these norms simply in virtue of being competent members of our communicative community. For example, it would be impermissible in ordinary circumstances for a mere acquaintance to attempt to gossip with me about my partner, or for my dissertation advisor to gossip with me about another student. This doesn't mean that these norms are never violated, or even that the implicit norms we abide by are ideal. Nonetheless, in the course of our communicative history, we were able to try and select for such norms. The new presence of theoretical discussions further constitute joint reflections on the norms that do and should govern the activity.

Subtweeting, we think, is *not* like gossiping in these respects (yet). Relatively speaking, our

life on Twitter has been short (and perhaps on the brink of extinction), so it wasn't possible to inherit rich knowledge of the norms that optimize the public benefits of the platform and its communicative nuances. Nor has it been subject to theoretical discussions regarding norm-articulation and construction. However, it is precisely these theoretical discussions and public conversations that must be initiated in order to identify the communicative effects of subtweeting and develop normative constraints on subtweeting. While, as we have argued in §4, subtweets—like gossip—might play important social roles, new norms ought to be developed to help users navigate the fraught linguistic and ethical challenges posed by them.

This is not to suggest that developing such norms will be easy. In fact, there will likely be various challenges. For example, in practice, it might prove difficult to distinguish between problematic *ad hominem* and important resistance subtweets. Hence, further work on subtweets would get clearer on when subtweeting is morally neutral or even a good thing, and help us identify norms for identifying and engaging with such tweets as well as for assessing the costs and benefits of them. Indeed, this should be part of a general project to clarify norms on social media more generally.[61] Although Twitter is the ideal ecosystem for this type of strategic speech, many of the insights we have offered are by no means limited to Twitter. As we have seen, there are variants of subtweeting on other social media platforms, particularly text-based ones. Moreover, any future platform that shares the features of Twitter highlighted will likely generate a variety of subtweet.[62]

At its best, Twitter is a medium of tremendous value. Countless users have benefited from its potential for building community, information dissemination, political mobilization, creativity, and more. By reflecting and developing communicative norms governing the new communicative practices afforded by the platform, and incorporating design elements that enable the realization of these norms, we might be able to optimize for its positive uses. In the meantime, we suggest that social media users be more mindful both of how they detect and engage with subtweets (and certainly about when they contribute to them). They may also be obligated to gather more evidence or information before positively engaging with a subtweet. Or, of course, we can always log off.

61. Cf. Rini (2017) on clarifying norms on social media testimony generally (e.g. "A retweet is not an endorsement").

62. In addition, we noted variants of subtweeting that appear offline in fn. 21, where deniability and inspecificity function in similar ways as discussed here. These phenomena are worthy of further exploration.

## Works Cited


Adkins, K. 2017. *Gossip, Epistemology and Power: Knowledge Underground*. Palgrave Macmillan.

Alfano, M. and Robinson, B. 2017. Gossip as a Burdened Virtue. *Ethical Theory and Moral Practice*, 20(3):473–487.

Alicart, H., Cucurell, D., and Marco-Pallarés, J. 2020. Gossip Information Increases Reward-Related Oscillatory Activity. *NeuroImage*, 210:116520.

Berger, J. and Milkman, K. L. 2012. What Makes Online Content Viral? *Journal of Marketing Research*, 49(2):192–205.

Berlyne, D. E. 1954. A Theory of Human Curiosity. *British journal of psychology*, 45(3):180.

Berstler, S. 2019. What's the Good of Language? On the Moral Distinction Between Lying and Misleading. *Ethics*, 130(1):5–31.

Bogost, I. 2015. Introducing the Supertweet. *The Atlantic*. https://www.theatlantic.com/technology/archive/2015/01/introducing-the-supertweet/384730/.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., and Bavel, J. J. V. 2017. Emotion Shapes the Diffusion of Moralized Content in Social Networks. *Proceedings of the National Academy of Sciences*, 114(28):7313–7318.

Bräuer, F. 2023. Statistics as Figleaves. *Topoi*, 42(2):433–443.

Camp, E. 2018. Insinuation, Common Ground, and the Conversational Record. In Daniel Fogal, Daniel W. Harris, M. M. (ed), *New Work on Speech Acts*. Oxford University Press, Oxford.

Cohen, T. 1999. *Jokes: Philosophical Thoughts on Joking Matters*. University of Chicago Press.

Dinges, A. and Zakkou, J. forthcoming. On Deniability. *Mind*.

Edwards, A. and Harris, C. J. 2016. To Tweet or 'Subtweet'?: Impacts of Social Networking Post Directness and Valence on Interpersonal Impressions. *Computers in Human Behavior*, 63:304–310.

Feinberg, J. 1987. *Harm to Others*, vol. 1. Oxford University Press.

Fricker, E. 2012. Stating and Insinuating. *Aristotelian Society Supplementary Volume*, 86(1):61–94.

Fritts, M. and Cabrera, F. 2022a. Fake News and Epistemic Vice: Combating a Uniquely Noxious Market. *Journal of the American Philosophical Association*, 8(3):1–22.

Fritts, M. and Cabrera, F. 2022b. Online Misinformation and 'Phantom Patterns': Epistemic Exploitation in the Era of Big Data. *Southern Journal of Philosophy*, 60(1):57–87.

Frost-Arnold, K. 2021. The Epistemic Dangers of Context Collapse. *Applied epistemology*, pp. 437.

Goffman, E. 1955. On Face-Work: An Analysis of Ritual Elements in Social Interaction. *Psychiatry*, 18(3):213–231.

Goldberg, S. C. 2021. The Promise and Pitfalls of Online ?Conversations? *Royal Institute of Philosophy Supplement*, 89:177–193.

Grice, H. P. 1975. Logic and Conversation. In *Speech acts*, pp. 41–58. Brill.

Guercio, N. L. and Caso, R. 2022. An Account of Overt Intentional Dogwhistling. *Synthese*, 200(3):1–32.

Habgood-Coote, J. 2019. Stop Talking About Fake News! *Inquiry: An Interdisciplinary Journal of Philosophy*, 62(9-10):1033–1065.

Haney-López, I. 2014. *Dog Whistle Politics: How Coded Racial Appeals Have Reinvented Racism and Wrecked the Middle Class*. Oxford University Press.

Henderson, R. and McCready, E. 2018. How Dogwhistles Work. In Arai, S., Kojima, K., Mineshima, K., Bekki, D., Satoh, K., and Ohta, Y. (eds), *New Frontiers in Artificial Intelligence*, pp. 231–240, Cham. Springer International Publishing.

Hunt, E. 2017. Vaguebooking? Subtweeting? Supertweeting? Why Can't We Just Say What We Mean Online? *The Guardian*. https://www.theguardian.com/culture/2017/may/26/vaguebooking-subtweeting-supertweeting-why-cant-we-just-say-what-we-mean-online.

Khoo, J. 2017. Code Words in Political Discourse. *Philosophical Topics*, 45(2):33–64.

Lee, J. J. and Pinker, S. 2010. Rationales for Indirect Speech: The Theory of the Strategic Speaker. *Psychological Review*, 117(3):785–807.

Litt, E. 2012. Knock, Knock. Who's There? The Imagined Audience. *Journal of broadcasting & electronic media*, 56(3):330–345.

Litt, E. and Hargittai, E. 2016. The Imagined Audience on Social Network Sites. *Social Media Society*, 2(1):205630511663348.

Marsili, N. 2021. Retweeting: Its Linguistic and Epistemic Value. *Synthese*, 198:10457–10483.

Marsili, N. and Löhr, G. forthcoming. Saying, Commitment, and the Lying–Misleading Distinction. *Journal of Philosophy*.

Marwick, A. E. and boyd, d. 2011. I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media & Society*, 13(1):114–133.

Marwick, A. E. and boyd, d. 2014. Networked Privacy: How Teenagers Negotiate Context in Social Media. *New Media & Society*, 16(7):1051–1067.

Mazzarella, D. 2021. "I Didn't Mean to Suggest Anything Like That!": Deniability and Context Reconstruction. *Mind & Language*.

McLuhan, M. 1994. *Understanding Media: The Extensions of Man*. MIT press.

Mellema, G. 2008. Professional Ethics and Complicity in Wrongdoing. *Journal of Markets & Morality*, 11(1).

Mendelberg, T. 2017. *The Race Card*. Princeton University Press.

Nester, D. 2015. In Defense Of "Vaguebooking". *BuzzFeed News*. https://www.buzzfeednews.com/article/danielnester/in-defense-of-vaguebooking.

Nguyen, C. T. 2021a. How Twitter Gamifies Communication. In Lackey, J. (ed), *Applied Epistemology*, pp. 410–436. Oxford University Press, Oxford.

Nguyen, C. T. 2021b. Twitter, the Intimacy Machine. *The Raven: A Magazine of Philosophy*. https://ravenmagazine.org/magazine/twitter-the-intimacy-machine/.

Parkinson, H. J. 2014. Subtweeting: What is it, and How to do it Well. *Vox*. https://www.theguardian.com/technology/blog/2014/jul/23/subtweeting-what-is-it-and-how-to-do-it-well.

Peet, A. 2015. Testimony, Pragmatics, and Plausible Deniability. *Episteme*, 12(1):29–51.

Peng, X., Li, Y., Wang, P., Mo, L., and Chen, Q. 2015. The Ugly Truth: Negative Gossip About Celebrities and Positive Gossip About Self Entertain People in Different Ways. *Social Neuroscience*, 10(3):320–336.

Reinsberg, H. 2013. What is a Subtweet? *BuzzFeed News*. https://www.buzzfeed.com/hillaryreinsberg/what-is-a-subtweet.

Rini, R. 2017. Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal*, 27(S2):43–64.

Rose, P. 1984. *Parallel Lives: Five Victorian Marriages*. Vintage.

Saul, J. 2018. Dogwhistles, Political Manipulation, and Philosophy of Language. In Daniel Fogal, Daniel W. Harris, M. M. (ed), *New Work on Speech Acts*. Oxford University Press, Oxford.

Saul, J. M. 2002. Speaker Meaning, What is Said, and What is Implicated. *Noûs*, 36(2):228–248.

Saul, J. M. 2012. *Lying, Misleading, and What is Said: An Exploration in Philosophy of Language and in Ethics*. Oxford University Press.

Saul, J. M. 2017. Racial Figleaves, the Shifting Boundaries of the Permissible, and the Rise of Donald Trump. *Philosophical Topics*, 45(2):97–116.

Segal-Gould, N. L. 2018. Don't Take This Personally: Sentiment Analysis for Identification of "Subtweeting" on Twitter. *Senior Projects Spring 2018*, 244.

Sharron and Abraham, J. 2015. The Role of Curiosity in Making Up Digital Content Promoting Cultural Heritage. *Procedia - Social and Behavioral Sciences*, 184:259–265.

Strong, C. 2015. Private Lives? https://www.mrs.org.uk/pdf/private%20lives.pdf.

Thompson, C. 2008. Brave New World of Digital Intimacy. *New York Times*. https://www.nytimes.com/2008/09/07/magazine/07awareness-t.html.

Tufekci, Z. 2014. Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. In *Eighth International AAAI Conference on Weblogs and Social Media*.

Tufte, E. 2003a. PowerPoint is Evil. *Wired Magazine, Septembre*, 10(11):9–11.

Tufte, E. R. 1990. *Envisioning Information*, vol. 126. Graphics press Cheshire, CT.

Tufte, E. R. 2003b. *The Cognitive Style of PowerPoint*, vol. 2006. Graphics Press Cheshire, CT.

van Niekerk, J. 2008. The Virtue of Gossip. *South African Journal of Philosophy*, 27(4):400–412.

Viebahn, E. 2021. The Lying–Misleading Distinction: A Commitment-Based Approach. *Journal of Philosophy*, 118(6):289–319.

Wiegmann, A., Willemsen, P., and Meibauer, J. 2022. Lying, Deceptive Implicatures, and Commitment. *Ergo*, 8.

Witten, K. 2014. Dogwhistle Politics: The New Pitch of an Old Narrative. *Unpublished Manuscript*.

Yablo, S. 2014. *Aboutness*. Princeton University Press.