

The Many-to-Many Model

Communication, Attention, and Trust in Online Conversations

Marcin Lewiński

NOVA Institute of Philosophy and Department of Communication, NOVA
University Lisbon

Catarina Dutilh Novaes

Department of Philosophy, Vrije Universiteit Amsterdam

1. Introduction

In recent decades, conversations supported by internet technologies, in particular on social media platforms, have become extremely pervasive. Are these conversations fundamentally different from offline, typically face-to-face conversations? Do they require different, novel conceptual models to be studied? Or perhaps an examination of conversations *online* makes us reconsider our ideas of what *conversations* in general are like in the first place? These are some of the questions we address in this contribution.

Any philosophical study of conversation and communication happens against the background of a conceptual model determining what conversation/communication is, can be, or ought to be. Traditionally, philosophical inquiry has revolved around two such models. The first is the *one-to-one* model of conversation, epitomised in Socratic dialogues, theorised in Aristotle's dialectic (*Topics*), practised to perfection in medieval disputations, and treated today, across the board, as the normal, unmarked model of conversation in its philosophically most significant instances (by Austin, Grice, Lewis, and Stalnaker, among many others). Here, the Speaker and the Hearer¹ engage in an orderly give-and-take of a genuine exchange, where, ideally, their overt (and largely cooperative) communicative

¹ We have concerns pertaining to the ableist undertones of this terminology. But as we are critically discussing traditions where these terms are widely used, it seems we have to at least mention them (if not use them).

intentions are transmitted and read against substantive, mutually recognised common ground or publicly available conversational score. This, in turn, allows them to converge on an epistemically fruitful or practically prudent result.

The second model of philosophical interest is the *one-to-many* model of public speaking known from ancient oratory to twentieth-century broadcasting. Due to some of the key characteristics of this model—notably, the participation framework consisting of an active Speaker and a (largely) passive Audience—it has been treated with suspicion, from Plato (*Gorgias*, *Euthydemus*, *Phaedrus*) and Aristotle (*Rhetoric*) to today's analytic philosophers (Frankfurt, 2005; Khoo, 2021; Langton, 2018b; Saul, 2024; Stanley, 2015). The lack of genuine interactive engagement between Speaker and Audience stifles self-correction and serious critical probing aimed toward explicitly achieved mutual understanding. In turn, this facilitates the occurrence of familiar abuses of speech—lying, bullshitting, misleading, conveying sneaky presuppositions, introducing not-at-issue content and unreasonable emotional appeals, enacting oppressive norms, etc.—making them potentially more effective *qua* abuses. This is a danger identified and feared by philosophers in the communicative practices of ancient sophists and present-day propagandists alike.

But how do these two models (one-to-one and one-to-many) fare when it comes to explaining conversations online? Not very well, it seems. Online (groups of) speakers can interactively engage with other speakers (as individuals or groups), expanding criss-crossing and overlapping conversational threads with many known and unknown others. Active content generators and influencers interact with other active producers of content, but also with trolls, likers, lurkers, moderators, chatbots, and potentially with massive global audiences. Altogether, from the perspective of the two classic models, online conversations violate many key characteristics of a “normal” conversational exchange. Thus, what seems to be required to investigate online conversations fruitfully is a different model of conversation, something like a *many-to-many* model, where multiple senders and receivers interact, switching roles constantly and addressing various people and audiences simultaneously.

To account for the specific characteristics of online communication, we first briefly review key features of the two traditional models and then present our own many-to-many model of communication, which we believe is more appropriate to theorise online conversations. This model, as we further argue, provides a fruitful conceptual background to better understand a number of novel features of online conversations. An important component of our analysis is describing the “conversational ecology” of online environments: technological, structural, and design features of these environments seem to significantly alter conversational affordances when compared to offline environments. However, we do not claim that the many-to-many model is only relevant for online conversations; in fact, further reflection suggests that many-to-many conversations *also* occur in offline environments, although at a smaller scale without the technological affordances of online media. This fact remains so far under-appreciated within the one-to-one

and one-to-many conversational paradigms (but see Aakhus and Lewiński, 2011; Lewiński and Aakhus, 2023). One example of an offline many-to-many conversation might be a debate in a House of Representatives, which requires a suitable space and technical support (microphones), as well as precise procedural rules.

Building on this description, we show how attention and trust must be first secured for conversational engagement to happen at all in online environments. To this end, we apply the three-tiered model of epistemic exchange introduced in Dutilh Novaes (2020b) to explicate the mechanisms through which many-to-many conversations emerge and are shaped online. We then move on to discussing two paradigmatic examples of phenomena pertaining to online “conversational ecosystems”: gamification of communication and trolling. As we argue, while these two are common features of conversations online, they require something like the many-to-many model to be properly grasped.

Altogether, our strategy to face the novelty of online conversations is to offer a conceptual response, focused on developing a model suitable for these kinds of conversations, and to draw some of its philosophical implications. The model has the advantage of offering a unifying perspective under which many relevant phenomena can be subsumed. We thus hope that the model will prove fruitful for analysing a wide range of online phenomena that tend to be tackled in piecemeal, disconnected ways, thus providing a useful conceptual framework for future research.

2. Two Influential Models of Communication

Philosophical reflection on communication and conversation has revolved around two contrastive models: one-to-one dialogue and one-to-many public speaking.² Plato’s dialogues such as *Gorgias*, *Euthydemus*, and *Phaedrus*, are among the most famous early examples of analyses systematically contrasting these models, with a characteristic evaluative tinge. The careful back-and-forth interaction of Socratic dialogue, it is claimed, is conducive to self-corrective and philosophically valuable conversation. By contrast, long-winded unconstrained speeches performed by sophists before a largely passive and possibly gullible audience contain the germ of corrosive, fallacious argumentation. This Platonic imagery has, for centuries, shaped the way philosophers apprehend communication, both descriptively and normatively. But rather than tracing the history of this otherwise well-known distinction, our aim here is to extract the features of each of the models that are instrumental to better grasping the specificities of online communication.

² Not surprisingly, communication scholars have introduced a variety of “communication models,” such as the Shannon–Weaver model, the Lasswell model, the Osgood–Schramm model, etc. (see McQuail and Windahl, 1993). While each of these models differs in features that are significant to communication scholars (e.g. the role of channel and noise in communication transmission, the functions of feedback) on the conceptual level we operate here, they are all variations of either the one-to-one or one-to-many model.

2.1 The one-to-one model

The one-to-one model captures the basic features of unmediated, face-to-face interaction—arguably, the most naturally experienced form of communication in any oral culture. It was extolled by Plato both via the animated renderings of dialogical mastery by his teacher Socrates and via conscious theorising.

In this model, communication amounts to a *synchronous* and *interactive* encounter between (typically) two interlocutors. The two parties communicate via an intentional, coordinated give-and-take of information, relying on resources such as mutually shared conventional language and non-verbal signals. These coordinated exchanges can also be used for purposes such as argumentation and debate, in which case the roles of Questioner and Answerer, Proponent and Opponent, or Protagonist and Antagonist have been implemented and theorised (see Dutilh Novaes, 2020a; Lewiński and Aakhus, 2023). Limited to the localised here-and-now, communication is essentially a *dyadic transaction* over which both parties wield *intentional control*.³ Their communicative contributions are largely dovetailed via standard, although infinitely expandable, sequences such as question–answer or argument–counterargument. These sequences are developed by interlocutors against the background of mutually recognisable *principles, maxims, rules, and conventions*. These, in turn, are undergirded by basic human capacities for *collective intentional action* and *practical reasoning* necessary for cooperation and mutual understanding. All the same, the capacities and rules also allow for various overt or covert *communicative strategies or projects* to emerge and be implemented by the interlocutors (e.g. Beaver and Stanley, 2019, 2023; Levinson, 2013). In this sense, the Speaker and the Hearer of the one-to-one model can be seen as players in a dyadic (language) game, not unlike chess or checkers.

Importantly, Speakers have *privileged control* over their communicative intentions to convey meanings to the one Hearer they address.⁴ Via self-corrective chains of questions and answers, they can reclaim and repair the hearer's derivation of speaker meaning and the force of their communicative contributions (e.g. Dingemanse et al., 2015). This intentional control is possible precisely because of the mutual availability of substantive common ground as described above

³ We say “essentially,” as the role of the onlooking audience has been recognised since antiquity: Socrates and his interlocutors typically debated in a public space, surrounded by disciples and other onlookers. Yet, the focus has been squarely on the two sparring partners, as it were, while the remaining parties have been largely relegated to the role of contextual paraphernalia.

⁴ We treat this conversational control in terms of a design feature of some idealised model of one-to-one dialogues where it is broadly “afforded.” However, it is by no means guaranteed, especially in the context of real-life “asymmetrical conversations,” where power imbalances prevent one of the interlocutors from performing or protesting certain speech acts (see Bianchi, 2019).

(including joint attention, shared goals, background knowledge, and mutual trust), which is continuously recalibrated.⁵

2.2 The one-to-many model

By contrast, the one-to-many model captures the experience of communicating to a group of people. Originally, it covers two fundamental phenomena: that of an orator speaking publicly before an audience in oral communication and that of an author communicating with their readers via written communication. (Radio and television broadcasting are, of course, further extensions of this form of communication.) Plato famously criticised both instantiations of this model, notably writing, in his dialogues (*Gorgias*, *Phaedrus*) and letters (*Letters II* and *VII*). One fundamental criticism is the lack of interactive engagement of the audience with the speaker/writer. The virtuous feedback loop between interlocutors is thus broken, jeopardising on-the-fly correction and mutual understanding. This is further exacerbated in the case of writing by the detachment of that which is communicated from the communicators: A text cannot speak for the author when queried by readers, thus allowing for unwarranted (misunderstood, superficial, deliberately distorted) uses, reuses, and abuses of the text.

The one-to-many model thus conceives of communication as largely *non-interactive* and, in the case of writing, also *asynchronous*, as the text can be consumed at any time and place.⁶ As the “to-many” phrase explicitly indicates, on this model, communication is a *public, poly-adic affair*, where possibly heterogeneous groups of hearers/readers engage with the spoken or written discourse. Some of these groups might be intentionally addressed by the speaker, some not, including audiences unknown to and unanticipated by the speaker. This, as Plato already understood, results in the sender’s *loss of commanding intentional control* over the hearers’ or readers’ derivation of speaker meaning, which takes on a life of its own, as it were. As speakers and writers struggle to direct their communicative contributions to many others, they may need to resort to various strategies of *multiple-audience design* (Clark, 1992, 1996; Cooney et al., 2020; Sacks et al., 1974; Yoon and Brown-Schmidt, 2019). But they can also resort to more nefarious forms of communicative strategies, namely those that bypass reason in the name of successful crowd appeal. This is precisely the critique mounted against

⁵ A reader would recognise this description as an abstracted simplification of the ideas on conversation and communication familiar from the work of Austin (1962), Grice (1989), Levinson (2006), Lewis (1979), Stalnaker (2014), Tomasello (2008), and many others. We focus on the common core of these distinct approaches, while avoiding disputes over intriguing details that would take us too far afield for the current purposes.

⁶ Traditional writing, while always asynchronous, can also be one-to-one and interactive, as in (old-fashioned) letter exchanges. For us, this would be a variant of the one-to-one model.

the one-to-many model as employed by the ancient sophists and today's propagandists (Beaver and Stanley, 2019; Frankfurt, 2005; Langton, 2018b; Khoo, 2021; Saul, 2024; Stanley, 2015).

Much of the philosophical inquiry into communication has revolved around the dialectic between these two models—as in the standard philosophy versus rhetoric dispute. But both sides of the dispute seem to be somewhat unreflectively converging on the same basic assumption: namely that all forms of communication are, one way or another, derivative of these two basic models. Below, we argue that this assumption is unwarranted—and it is precisely the examination of conversations online that lets us lay this bare.

3. The Many-to-Many Model: Off- and On-line

While our brief description of the two influential models of communication is quite schematic, it provides a valuable contrastive background for our main focus: the many-to-many model. An important observation is that many-to-many communication—where groups of individuals address one another via interactive exchanges—has always been around. Plato's *Symposium* is a good example of such complex *polylogues*. But it is via the examination of conversations online that we can (a) retrospectively better perceive the existence of many-to-many communication even in offline communication and (b) better demarcate the features of many-to-many exchanges that are specific to online environments, including the novel technological affordances for conversations at a scale unavailable for offline exchanges.

Indeed, the affordances of typical offline media for communication—spoken language, asynchronous writing—as well as human cognitive limitations in terms of signal-processing impose upper bounds of *scale* to offline many-to-many conversations. Beyond a certain number of participants, a face-to-face conversation quickly becomes unfeasible, unless there are facilitating factors in place, such as a specific spatial design and precise rules for who speaks when (such as in debates in a House of Representatives). By contrast, the affordances of online environments (e.g. synchronous writing with near-immediate and archived feedback) allow for many-to-many conversations to involve a much larger number of participants, unconstrained by geographical distances and the fleeting existence of spoken language.

3.1 The many-to-many model: Offline barebones

Many of our daily communicative encounters—from informal family dinners to disputes on public transportation and formal contexts such as academic conference presentations—are neither simple, one-on-one dialogues nor one-to-many

speeches. Instead, they involve complex collective *poly-logues*, that is, discourses (λόγοι) between many (πολύ) (Lewiński, 2021; Lewiński and Aakhus, 2023). As such, they instantiate many-to-many modes of communication. Here, communication is largely *interactive* and *synchronous*, with real-time feedback characteristic of one-to-one dialogues. All the same, speakers communicate with a (possibly large) collections of hearers, similarly to the one-to-many situations. Yet, in contrast to public speaking, various third parties, bystanders, eavesdroppers, intended and unintended audiences, can easily tap into conversation. Given the potential complexity and scale of the exchanges—far exceeding the dovetailed *you speak–I speak–you speak* etc. patterns of a simple dyadic exchange—speakers wield *limited to no intentional control* over the direction and outcome of each conversation encounter. In particular, speakers have *limited to no control* over the derivation of the meanings they intend to convey by the multiple recipients of their speech acts. Unexpected events are likely to derail any attempts at strictly controlling or strategising one's communicative contributions directed to and from the many others (as anyone who has attempted to conduct a meaningful conversation on social media platforms surely knows . . .).

So conceived, the many-to-many model correctly captures and predicts many ordinary phenomena of communication (see especially Lewiński and Aakhus, 2023: ch. 4): much increased complexity of exchanges and often convoluted sequential patterns (multiple responses to a single speech act; parallel sub-discussions which might criss-cross, overlap, split and reunite, etc.); multiple-recipient design of speech acts; difficulties in gauging and establishing the common ground among all the participants; various forms of co-production of discourse that may lead to strategic coalition-building and majority–minority divisions; translation, mediation, and arbitration of communicative contributions by third parties; distributed or delegated responsibility for talking; problems in determining the relevance and completeness of exchanges. The list continues—and sociolinguists and pragmatists such as Clark (1992, 1996), Goffman (1981), and Levinson (1988) relish confronting the reality of such complex communication with the reductionist philosophical models we outlined above.

All the same, not all such complexities have been lost on philosophers. Egan (2009) considers complex, multi-audience communicative situations as a pretext to offer fixes to the meta-semantics of communication via the concepts of *audience sensitive contextualism* and *shotgun assertions* such as “Jesus loves you” when read and actualised by different “yous.” Sbisà (2013), Johnson (2019, 2023), and Lewiński (2021) turn their attention to *illocutionary pluralism*, the possibility of performing plural illocutionary forces to different hearers, with one and the same utterance: whether by the speaker's complex intention or by hearers' varied uptake. Similarly, McGowan (2023) discusses complex forms of communication involving multiple contexts of interpretation and multiple timelines and engaging overhearers and other agents in ways that challenge the simple dyadic model of scorekeeping in language games offered by Lewis (1979). Saul's (2024) study

of dogwhistles and Langton's (2018a) study of back-door speech acts also presuppose something like the many-to-many model, even if mostly implicitly (see Lewiński, 2021). Finally, epistemologists such as Zollman (2013) and O'Connor and Weatherall (2019) conceptualise phenomena of complex communication in terms of networks of senders and receivers and complex signalling games. Thus, we certainly do not take ourselves to be the first ones to recognise the importance of many-to-many communication. But here we seek to make explicit what often remains implicit in these discussions: the structural features that distinguish many-to-many conversations from the two influential conversational paradigms discussed above.

3.2 Many-to-many communication online

While many of the elements of online many-to-many communicative situations are present in their offline predecessors—and in theorising about them—as a set of features, they arguably give rise to a whole new phenomenon. What are these features?

As opposed to many offline siblings, conversations online present new and unique possibilities for intentional design of conditions—indeed, entire technologies—for communication (Aakhus, 2017; Aakhus and Jackson, 2005; Forestal, 2022; Paglieri, 2017). With simple design interventions—limiting the length of “posts,” organising discussions in thematic threads or simply chronologically, allowing only for likes or also dislikes, implementing immediate or delayed (i.e. synchronous vs asynchronous) feedback, etc.—the shape and outcome of conversations can be significantly altered (see Esau et al., 2017; Meredith, 2017). One of such “designable” features is the anonymity or pseudonymity that many of the online communication platforms allow. Notoriously hard to achieve in offline contexts, anonymity has been both lauded as the great status leveller in public sphere deliberations (see Jaidka et al., 2022) and as the great destroyer of accountability that erodes communicative norms and actively stimulates unwarranted and uncivil discourse (e.g. Goldberg, 2013). Either way, the possibility of anonymity and pseudonymity requires specific strategies to evaluate how worthy of engagement an interlocutor may be. While these phenomena are not entirely absent in offline conversations, they are arguably more widespread in online conversations, thanks to technological affordances.

Further, part of the design of online communication systems are hidden algorithms governing what we see first and how. It is well known that tech giants such as Google and Meta seek primarily to maximise the time that users spend online (so that they can resell attention to advertisers), and to this end, they employ algorithms that by and large determine what a person gets to see in her newsfeed (Forestal, 2022; Hindman, 2018). This means that, while we have some degree

of control over the sources and contents we are exposed to (e.g. whom we follow on social media), some conversations will be made salient to us by the algorithms, while others will be buried under all the “noise.” In online environments, the competition for attention and visibility is fierce and, to a large extent, determined by interests and goals that are orthogonal to facilitating fruitful conversations.

What is unique to most social media is that, by design, they afford and promote massive, geographically unconstrained, active participation of thousands of users. This potentially unlimited scale and interactivity is in stark contrast to more traditional offline contexts. In the latter, either the scale needs to remain small for many-to-many interactions to be truly interactive or massive participation prevails but only as passive participation of mass audiences of one-to-many broadcast media. Online, audiences can and do speak back (though there is no guarantee that they will be “heard”). In fact, the very concept of “audience” is not really appropriate. Conversations online offer a whole set of new communicative roles: from “posters” and “content creators” to “influencers,” “trolls,” “likers,” “lurkers,” “moderators,” “bots,” etc. While some of these roles might have approximate offline counterparts (think, for instance, of hecklers for trolls and head-nodders for likers), anonymous online contexts afford their striking crystallisation, thus making us more acutely aware of their functioning. Moreover, none of these roles correspond in any straightforward way to the traditional roles of speakers, hearers, and audiences. An upshot of this many-to-many conversational structure is that people typically have a gigantic range of possible interlocutors to engage with, thus constantly having to make choices (more than in offline conversations, it seems, where material constraints limit the range of options). Consequently, senders of signals are competing with each other for the attention and engagement of receivers, which means that signals sent may fail to have the desired uptake or even any uptake at all (something that is presumably less likely to occur in one-to-one and one-to-many conversations).

Adding to the complexity is also the fact that online communication is increasingly multimodal: next to traditional written texts, we rely more and more on emojis, memes, images, gifs, voice messages, videos (Herring, 2019; Jenkins et al., 2013). The incredible success of platforms such as Instagram and TikTok attests to this.

Similar to offline many-to-many conversations, online communication has attracted little philosophical attention, which is, however, gradually picking up. There is certainly much catching up to do: linguists and communication scholars have dedicated concerted effort to analysing online communication for three decades now, making computer-mediated communication (CMC) a field of study in its own right (Herring, 2019). In analogy with such linguistic studies, there is work that reflects on curious characteristics and patterns of online conversations, which then scale up to broader, and philosophically worrisome, communicative dynamics. Lewiński (2010) examined how contributors to online

discussion forums team up to collectively criticise someone else's argumentation. Such collectives can be impromptu, without having one consistent position (as in Lewiński's study), but they can also gradually form a coherent group with an increasingly entrenched position and a deliberate policy to exclude dissenting voices. "Epistemic bubbles," which filter out uncomfortable content, initially perhaps unconsciously, easily turn into self-conscious collective "echo chambers" where opposition is actively eliminated (Edwards, 2013; Nguyen, 2020), based on their non-conformity with the "identity-defining beliefs" of the group (Williams, 2022). This process is then reflected in the well-known phenomenon of online polarisation (Sunstein, 2007).

In a similar vein, McDonald (2021) analysed the communicative functions of "likes" on platforms such as Facebook, and Marsili (2021) those of "re-tweets" and other re-uses of online content. They both point to "a corrosive motivation to online communication" (McDonald, 2021: 335) driven by the fact that likes, re-shares, and numbers of followers generate a quantifiable tally of one's popularity or social capital online. This makes the thrill of producing likeable and re-tweetable messages more prominent than that of generating epistemically valuable content conducive to socially significant deliberation. This *gamification* of communication (Nguyen, 2021) is one of the two exemplary phenomena we analyse below.

Two further features specific to online many-to-many conversations stand out as consequences of its complex participation framework and technological affordances. First, "speakers" typically directly interact with many others, which calls for different, possibly confusing, strategies of production and interpretation (Lewiński, 2021). Trolling, the second phenomenon we discuss below, is a paradigmatic instance of this. Second, interacting with unknown others and via hidden algorithms, "speakers" are deprived of the sort of intentional control over communicative contributions that forms the basic principle of the traditional "logic of conversation." Below, we call attention to this challenge and begin drawing some of its philosophical implications.

4. Attention, Trust, and Engagement Online

In order to formulate a more detailed analysis of online conversations following the many-to-many model, we propose to take a closer look at the structure of online communicative ecosystems. To do so, we expand and adapt the three-tiered model of epistemic exchange (TTEX for short) that one of us has been developing in recent work (Dutilh Novaes, 2020b; Ivani and Dutilh Novaes, 2022; Dutilh Novaes, 2023). TTEX highlights the importance of attention and trust in epistemic and communicative exchanges; we will see that these two factors acquire specific contours in online environments.

4.1 The three-tiered model of epistemic exchange

TTEX was inspired by a framework known as social exchange theory (SET) (Dutilh Novaes, 2020b). SET was developed by sociologists and social psychologists in the mid-twentieth century to explain human social behaviour in terms of processes of exchanges between parties, involving costs and rewards, and against the background of social networks and power structures (Cook, 2013). It was influenced by research in economics (rational choice theory), psychology (behaviourism), and anthropological work by Malinowski, Mauss, and Lévi-Strauss (Cook, 2013).

SET is an empirically robust framework, which has been used to investigate a wide range of social phenomena such as romantic relationships, business interactions, and trust in public institutions, among many others. In particular, and relevant for our purposes, it has been extensively used to investigate interpersonal communication (Roloff, 2015). The SET models are neither purely descriptive—as they rely on certain normative assumptions such as that agents seek to maximise rewards and minimise costs, while also adhering to principles of reciprocity—nor purely normative, given that they incorporate experimental findings as well as extensive observational data. Moreover, SET combines a first-person perspective, which explains and predicts choices that individuals make between different potential exchange partners, with a third-person perspective, which focuses on structural features of these exchange networks.

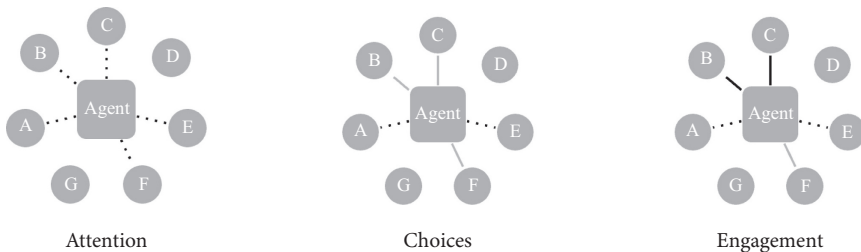
TTEX adapts the SET findings pertaining to exchanges more generally to exchanges that are specifically epistemic, and here we extend the main features of TTEX to *communicative* exchanges. This leads to distinguishing three main levels in processes of communicative exchanges:

1. **Attention/exposure.** The first level consists in the networks determining whether people are potential exchange partners, given the relevant opportunity structures for epistemic exchange. Who is in a position to attract the attention of others? It may be that potential lines of communication are cut, for example in the case of censorship (Dutilh Novaes and de Ridder, 2021) or echo chambers (Nguyen, 2020). But it may also be that very many signals are being broadcast simultaneously, meaning that several different sources are competing for the receiver's attention (Gershberg and Illing, 2022), in a so-called “attention economy” (Franck, 2019).
2. **Choosing whom to engage with.** The next level comprises the choices that agents make against the background of possibilities for exchange, as determined by the relevant opportunity structures. Typically, there will be a number of options for a given agent, but given limitations of time, attention, and resources, contrastive choices will have to be made. Among those sources that have caught my initial attention, whom do I further engage with

(if at all)? At this point, considerations of trustworthiness (Hawley, 2019) come into play,⁷ as well as the perceived value of the content being offered by different potential exchange partners.

3. **Engagement with content.** It is only at a third stage that engagement with content properly speaking occurs; this is when the actual communicative exchange in fact takes place. At this point, the receiver will engage with the content in question, seeking to understand its substance. This may lead to a change in view for the receiver (though, of course, even at this stage the receiver may still balk at revising her beliefs) and to an update of the common ground for the agents involved in a particular interaction.

The diagrams below represent the three tiers. (For simplicity, they depict a main agent and other agents around her, even though the model focuses on complex networks of agents.)



- **Attention:** Agent does not “see” sources D and G; the other sources catch her attention (dotted lines).
- **Contrastive choices:** Agent deems B, C, and F as worth exchanging with (grey lines), but not A and E.
- **Engagement:** Agent eventually engages substantively with B and C (black lines), but not with F.

A concrete example will help illustrate the model. Consider the range of newspapers that a person can read to stay informed. Of all the existing newspapers, a given person is aware of the existence of only a subset of them (tier 1). Among those that are within the person’s attentional network, some she will consider of low quality or unreliable; these are then discarded as potential sources of information (tier 2).

⁷ Trustworthiness, which we understand as comprising an *epistemic* component (competence) and a *moral* component (benevolence, sincerity) (Dutilh Novaes, 2020b), is particularly relevant in communication that involves information exchange. Of course, many communicative engagements have other purposes (e.g. jocular exchanges), in which case trustworthiness thus understood may be less relevant. But even in such cases, we seek interlocutors whom we can trust not to abuse the purposes of the communication (e.g. jokes that quickly derail into insults) (Herbert and Kukla, 2016).

Finally, the agent will typically read only a subset of those that are considered sufficiently reliable/interesting, given limitations of time and cognitive resources (tier 3).

TTEX is particularly useful to analyse many-to-many conversational structures because it underscores the fact that actual communicative exchanges are premised on there being potential communication channels, as determined by the topology of the relevant attentional networks (tier 1), and on agents making contrastive choices between the different possible interlocutors in the relevant networks (tier 2). Both these factors remain undertheorised in the one-to-one model, where the two interlocutors have already secured mutual attention, and have already chosen to engage in communicative exchange with each other. They are also insufficiently highlighted in the one-to-many model, where one sender commands the attention of many (though attracting attention in such cases is not a given, as the sender must “lure” the audience) and the receivers are by and large conceived as passive conversational partners who do not make contrastive choices (beyond the initial choice of being members of this audience or not). Both models take as given the conditions of possibility for communicative exchanges to occur at all, whereas TTEX problematises and explains these conditions of possibility.

4.2 Attention

Online conversations differ significantly from offline conversations at each of the three tiers of TTEX. Regarding tier 1, let us start with the observation that attention is characterised by being contrastive—it involves focusing or concentrating on some sources/objects in place of others—and by involving some degree of voluntary control and simultaneously being susceptible to external influences (Smith and Archer, 2020). Online conversations are not limited by spatial and geographical constraints, so in theory at least, the possibilities for engaging in multiple conversations are endless. (This is, of course, truer in platforms that invite unrestricted connections, such as Twitter, where the default privacy setting is “public,” than in platforms such as Facebook, which more easily allow for content to be accessible only to predetermined circles of receivers.⁸) But since attention is contrastive, I cannot pay attention to, much less actively participate in, *every* ongoing conversation. In one-to-one conversations, by explicitly *addressing* a potential interlocutor, I direct her attention towards me to initiate the conversation, whereas

⁸ An interesting phenomenon on platforms such as Facebook is when a person’s different offline social circles come to “collide” in conversations prompted by the person’s posts. All of a sudden, your conservative uncle is engaging in heated discussion with your Black Lives Matter activist friend, a “conversation” that is very unlikely to take place in offline environments simply because these two people will most likely never meet in person. In online conversations, A being a potential conversational partner of B and C often leads to B and C also becoming potential conversational partners of each other. See also the concept of “context collapse” (Frost-Arnold, 2021; McDonald, 2025).

in online environments, this initial step is less systematic. (But note that one can “tag” people in social media to attract their attention, or use the reply/comment function to react to a specific statement.) Indeed, different potential conversational partners are simultaneously sending signals and competing for attention, a resource that is not evenly distributed (Hindman, 2018). Senders with large numbers of “followers”—celebrities, influencers—enjoy much more visibility and thus much more power to command attention. The internet has not reduced the cost of reaching audiences—it has merely shifted who pays, and how (Hindman, 2018).⁹

Moreover, as already noted, we have limited control over the different contents and potential partners we are exposed to online, given the role played by algorithms determining what appears in our “newsfeed.” We may carefully curate whom we follow on various platforms (the dimension of voluntary control), but platforms may easily override our efforts to manage our attentional regime ourselves (the dimension of external influences). Since the main goal of these algorithms is primarily to maximise the time people spend online and to generate clicks (“clickbaits”), not to facilitate the careful exchange of information, these attentional structures are often far from optimal for certain kinds of conversations (while perhaps supporting other kinds of conversations, such as idle chit-chat to pass the time).

In sum, attentional networks that are in part shaped by individual choices and in part determined by design features of platforms both enable and constrain potential conversational channels. Thus, while in theory everyone can talk to anyone online, in practice, the range of potential conversations is determined by these attentional networks and by the fact that attention is a limited resource that entails the need to be selective.

4.3 Choices and trust

Let us now turn to tier 2 phenomena, pertaining to the choices we make among potential interlocutors with whom to engage. As most of our potential conversation partners online are unknown to us in real life (and many present themselves anonymously), our assessments of whether we should engage with them must follow different strategies than in offline contexts.¹⁰ Should I engage with this Twitter¹¹ user who relentlessly comments on my tweets? Is he or she engaging

⁹ Attracting and commanding attention online is also a crucial aspect of systematic, sustained campaigns of disinformation, with troll factories (Pomerantsev, 2019) and “flooding the zone” strategies (Starr, 2020), but also of activism and social justice campaigns (Blevins et al., 2019).

¹⁰ Of course, there is typically some degree of overlap between our online and offline conversational partners. In particular on social media platforms such as Facebook, online networks (partially) reflect offline networks, as people become “Facebook-friends” after having met in person (or the other way round), and have their existing friends and family as social media contacts.

¹¹ This chapter was first written before the transition from Twitter to X. What we say about Twitter here pertains to the platform as it was before the transition/rebranding in July 2023, hence our consistent use of the platform’s previous name. As is well known, beside the name, many design features were altered as well, fundamentally changing the “flow” of conversations in the new platform.

in good faith, or am I dealing with a troll? (Or maybe even a bot?) We are constantly assessing the prospects for fruitful interaction with the multiple potential interlocutors that present themselves to us, as not all conversations are worth having. In offline conversations, there is typically the perception that it is simply a matter of politeness to respond to someone who addresses us. Online however, it is often inevitable (and highly advisable) to make choices as to whether to pursue or else to discontinue engagement with some interlocutors. (Needless to say, many online conversations end up being noxious or, at the very least, rather frustrating.) Indeed, it is sensible to disengage from conversations if it becomes apparent that the interlocutors involved are not worthy of our time and attention, or else that they are trying to mislead or manipulate us (i.e. are not trustworthy).

This being said, online environments do also allow for conversations between “strangers,” which diachronically may give rise to trust and mutual respect. A telling real-life example is the case of Megan Phelps-Roper, originally a member of the Westboro Baptist Church (a hyper-Calvinist congregation often described as a hate group). Westboro members are known for engaging in inflammatory homophobic pickets, as well as hate speech against atheists, Jews, Muslims, transgender people, and numerous Christian denominations. Megan joined Twitter initially to spread the Westboro world view. However, through engagement in conversations on Twitter with people who were initially perfect strangers to her and in fact held identities that the Westboro worldview considered inferior (Jews, homosexuals), she came to respect and seriously consider their critiques of her positions (Chen, 2015). Megan embarked on a long and painful process of questioning everything she had been brought up to believe (Dutilh Novaes, 2023). This process led to a full “conversion”; she eventually made the consequential decision to leave the church and began connecting again, including in offline environments, with some of her Twitter contacts.¹² (She even went on to marry one of them.)

Less life-transforming but still fruitful conversations can and do also occur online among people who share similar interests, such as communities of academics, fans, people belonging to marginalised groups, etc. Some platforms have features that allow for these groups to coalesce, such as Facebook groups. Within those communities, a sense of belonging and familiarity may arise through these ongoing conversations, which can be constructive but may also give rise to problematic echo chambers where, as mentioned above, contesting voices are systematically silenced and excluded (Nguyen, 2020).¹³

The bottom line is that, as originally conceived and designed, the internet presupposes truthful communication between trustworthy interlocutors; recall that the World Wide Web was developed by researchers working at the European

¹² Of course, this is the same process that may lead perfectly reasonable people to come to embrace outlandish conspiracy theories by forging connections within online communities. The potential for radicalisation of internet platforms is an ongoing source of concern (Koehler, 2014).

¹³ See e.g. the role of Facebook groups in the spread of disinformation about vaccines (Schmidt et. al., 2018).

Organization for Nuclear Research (CERN), to be a “pool of human knowledge” (Berners-Lee et al., 1994). As such, platforms supporting online interactions are typically poorly equipped to single out and flag conversations where trust and trustworthiness are not to be taken for granted. Recent technologies, such as blockchain, seek to remedy what some describe as “a problem of trust” on the internet, but for now, they still seem to be poor substitutes for tried and tested (but by no means infallible) offline mechanisms to ensure trustworthiness in communication such as strong social ties and prior familiarity. (Of course, there is also much deception offline, and we are notoriously bad at identifying lies (Levine, 2019).)

4.4 Engagement

Finally, we address tier 3 phenomena, when actual engagement occurs, after suitable relations of attention have been established and interlocutors have chosen to directly engage with each other. Design features of different online platforms lead to different kinds of conversations, involving more or fewer participants and different possibilities for embedded conversations, among other variables. Indeed, while on a platform such as Facebook comments always remain visibly subsumed to the original post, on Twitter, sub-conversations may acquire a life of their own. Each new tweet, whether a reply or a new tweet, is a communicative act at roughly the same “ontological level” from the point of view of the platform’s design.¹⁴ Such multi-layered, interwoven conversations are perfect examples of what we mean by many-to-many communication.

Moreover, in these conversations, the intended audience is *both* the specific explicit interlocutor(s) we are engaging with and a presumed broader audience, which is following the conversation more passively (Marcoccia, 2004; Lewiński, 2013). There are analogous offline situations where interlocutors are explicitly addressing each other but simultaneously also performing for an audience (e.g. in a political debate). But while such situations are arguably the exception rather than the norm in offline conversations, the opposite is the case in online conversations, which are both public *and* interactive (as highlighted by the many-to-many model). One of the upshots of such conversations involving multiple agents, which may then further split into sub-conversations, is that common ground is much more volatile and elusive in these situations than in one-to-one (offline)

¹⁴ One design feature of Twitter, namely the limit on the number of characters allowed per tweet, was initially the object of much exasperation among its users. Twitter later increased the maximum number of characters from 140 to 280 and introduced the *threads* feature, which in practice makes the limited number of characters much less significant (as one can post threads that are as long as one wishes, chaining tweets that are easily read in sequence but can be replied to individually). Another significant feature we do not have the space to discuss here in any detail is Twitter’s hashtags; their complex communicative functions (Zappavigna, 2015) are amenable to analysis within the many-to-many model and the TTEX approach.

conversations, as also captured by the notion of context collapse (Frost-Arnold, 2021). This is, in part, what makes online conversations so challenging and more likely to derail.

Furthermore, there are different kinds of communicative interventions online, which range from explicitly addressing someone with a reply to more subtle but equally significant communicative actions such as likes (McDonald, 2021), re-shares (Marsili, 2021), and “lurking” (absorbing content without signalling one’s presence in any obvious way; see Sun et al., 2014). Given that such interventions are easily quantifiable in terms of the numbers of likes, re-shares, and followers, online conversations lend themselves to a rather explicit score-keeping that is typically absent in offline conversations (more on score-keeping and gamification below).

These observations give rise to a number of normative questions, such as: Which online conversations are worth having? What are our communicative obligations vis-à-vis each other in online environments? How does one exercise communicative agency in these environments, especially in view of the role of algorithms and the power of tech corporations in shaping online conversations? Who are the presumed audiences for online discourse in different instantiations? These are questions that we cannot fully address here, but which can be fruitfully approached from the TTEX perspective.

5. Examples

To further illustrate the fruitfulness of our analysis, in this section we discuss two interesting phenomena pertaining to online communication: the gamification of online conversation and trolling.

5.1 Gamified conversations

The idea that linguistic communication is something like a game is familiar from the work of the late Wittgenstein, with the notion of “language games” (*Philosophical Investigations*, 2001), from Lewis’ (1979) work on conversational score-keeping (see also Langton, 2015; McGowan, 2019; Witek, 2021) and, more recently, from Brandom’s description of the “game of giving and asking for reasons” (Brandom, 1994). But the senses in which communication and conversation are game-like have acquired new contours in online environments, which are worth exploring.

In a particularly insightful article, C. Thi Nguyen (2021) approaches the topic of online conversations equipped with the concept of *gamification*. This concept has a number of enthusiastic supporters, most famously game designer Jane McGonigal, who wrote the tellingly titled book *Reality Is Broken: How Games*

Make Us Better and How They Can Change the World (McGonigal, 2011). For those enthusiasts, gamification is the solution to the “dreadfulness” of everyday life, which is filled with dull activities such as work, education, and chores. If only we could make these activities more fun . . . The thought is that, by importing key design features from modern gaming, we *can* in fact make these activities more fun—and as a consequence, life more bearable. Thanks to the powerful feedback and reward mechanisms available in games, otherwise dull activities are incentivised by the thrill of “point-scoring” (think of Duolingo as a way to learn new languages).

Despite being a fan of games himself, Nguyen has reservations toward the idea of turning all life activities into games. He thinks that communication, and online communication in particular, risks becoming thinned out and oversimplified if it is fully gamified. (Nguyen focuses on Twitter, but much of what he says applies *mutatis mutandis* to other online platforms.) In his own words,

Twitter shapes how we interact, who we interact with, and—perhaps most importantly—it suggests specific goals for those interactions. [. . .] Twitter shapes our goals for discourse by making conversation something like a game. Twitter scores our conversation. And it does so not in terms of our own particular and rich purposes for communication, but in terms of its own pre-loaded, painfully thin metrics: Likes, Retweets, and Follower counts.

(Nguyen, 2021: 410)

More specifically, “Twitter gamifies communication by offering immediate, vivid, and quantified evaluations of one’s conversational success” (Nguyen, 2021: 411). Scoring points becomes the goal of these conversations in and of itself, rather than the multifaceted and complex goals of human communication more generally. Indeed, thanks to its point-scoring system, Twitter is highly *addictive*. It is no coincidence that various social media platforms lean heavily on design strategies developed by the gambling industry,¹⁵ which seek to elicit small dopamine kicks at each like, re-share, or new follower.

Nguyen’s analysis is compelling and stands on its own: The thrill of scoring points offers by itself strong motivation for users to pursue likes, re-shares, and an increasing number of followers when engaging in online communication. However, the many-to-many model and TTEX offer further rationales as to why users want, and in fact *need*, to maximise these scores, at least if they want to be active participants in conversations and for their communicative actions to have uptake.

¹⁵ These connections have been somewhat dramatically, but by and large accurately, depicted in the Netflix docudrama *The Social Dilemma*.

The many-to-many model entails that there is no guarantee of becoming a respected, recognised participant in meaningful conversations. Given the multiple prospective interlocutors and audiences, a place at the table needs to be earned, as it were. And how does one become viewed as a valuable conversational partner in complex networks of possible interlocutors? *By scoring points*, which (presumably) directly measure one's prior conversational success and, in turn, make it more likely that future conversations (with old and new partners) will ensue.¹⁶

Moreover, TTEX highlights the fact that commanding attention (tier 1) and being viewed as a worthy conversational partner (tier 2) are conditions of possibility for conversations to happen at all. Without attention, one is not even *seen*, and without a certain degree of respect and perceived worth as a valuable interlocutor, one is *ignored* (even if seen). Amassing large numbers of followers prompts not only the thrill of “scoring points,” but also ensures larger visibility in a given network and thus more epistemic and communicative power (Archer et al., 2020). (So-called influencers typically engage in sustained strategies to increase their number of followers.) A similar reasoning applies to re-shares, which increase the visibility of specific communicative actions (and may, in turn, lead to an increase in number of followers).

Point-scoring on social media also affects tier-2 phenomena. One parameter that a social media user can rely on to ascertain whether they should further engage with a hitherto unknown user is whether they have “mutuals” and, more generally, whether they are followed by people whom the first user holds in sufficiently high esteem: this provides credentials to the hitherto unknown user. Granted, in offline environments similar mechanisms are in place (“this person is the friend of a friend”), but social connections are made explicit by the relations of following/being followed on social media platforms such as Twitter and Instagram (but not so on, for instance, Reddit).

Moreover, individual communicative actions (individual tweets) that amass a large number of “likes” receive an explicit seal of endorsement (re-shares are more ambiguous, as re-sharing can happen precisely to express disagreement), thus making it more likely that those following the conversation will be impacted by their content (be it in the form of further engagement or simply as “lurkers”). Diachronically, an interlocutor who is consistently “boosted” (with likes and re-shares) by other interlocutors whom we previously trust and appreciate will typically come to be seen as a valuable interlocutor as well.

In sum, the design features of Twitter and other social media platforms make it both *possible* to turn conversations into explicit games—thanks to the relevant

¹⁶ Another game-like strategy of securing conversational involvement is to skilfully engage in playful speech practices of (online) community insiders, for instance by using particular jargon or cultural references. Yet, contrary to manifest scoring, these practices tend to be somewhat occult, risking harmful misfires and the resulting “outgrouping” of an aspiring conversational partner (see Herbert and Kukla, 2016).

technological affordances—and *necessary* to maximise point-scoring, in view of the imperative to establish oneself as a sufficiently visible and worthy conversational partner (especially in circumstances where there are so many potential partners).

These game-like features of most social media communication alter the motivational architecture (incentive structure) of many-to-many interactions. Simple numerical scoring offers us, in the words of Nguyen, “a momentary experience of *value clarity*” (2021: 415), as the complex all-things-considered value judgments that we routinely resort to when communicating are reduced to powerful, objective, and immediately comparable quantified results. The problem Nguyen identifies is that our motivations shouldn’t be so reduced—or not so easily; online communication is still a complex, multidimensional human activity that happens in real life, not some invented gaming space.

This is reflected in how the language-game analogy is also a disanalogy. Wittgenstein referred to a language game as “the whole, consisting of language and the activities into which it is woven” (*Philosophical Investigations*, 2001, §. 7). As such, language games are complex, real activities, even if (just like *mere* games) they can be strategic and driven by the competitive *agon*. Accordingly, Lewis’ conversational score of a language game is not a numerical score but rather “an abstract data structure that tracks and represents publicly recognizable effects that speech acts have on the state of conversation” (Witek, 2021: 6212). These include presuppositions, standards of precision, points of reference for interpreting anaphoric expressions, felicity conditions of speech acts, and other mutually recognisable commitments of the “players.” Now, the problem is that this thick, qualitative tally of communicative commitments can only be successfully tracked for one-to-one, or otherwise small-scale communicative exchanges. Many-to-many activities quickly overload such complex scorekeeping and hence require something else to calibrate our attention, trust, and engagement at scale, even with the often available record of what has been said so far in an online conversation. Scoring in terms of points—likes, re-shares, and followers—is just one such design feature that ends up regulating how these conversations come to unfold.

5.2 Trolling

A common feature of online conversations recognised at least since the 1990s is trolling, that is, “mak[ing] a deliberately offensive or provocative online post with the aim of upsetting someone or eliciting an angry response from them” (*Oxford English Dictionary*, 2021). However, we contend that what exactly trolling is and how it functions can only be grasped against the conceptual background of the model of many-to-many online communication we outlined above.

Trolling, as reflected in its philosophical analysis (Barney, 2016; Cohen, 2017; Connolly, 2022; Ellefson: Chapter 8 in this volume; Morgan, 2023), relies on a complex communicative ontology. This ontology inescapably involves the troll, the target or victim of the trolling, and its complex audience. While terminology varies,¹⁷ let us distinguish between the *troll*, the perpetrator of trolling; the (often collective) *victim* that the trolling is about; the *outgroup audience* who is fooled by the troll; and the *ingroup audience* meant to recognise and enjoy the trolling. Importantly, the *outgroup audience* is often, but not necessarily, identical to the *victim*. This happens, for instance, when people who identify as migrants respond to trolling about migrants. Yet, a troll can also attack migrants while provoking an (often collective) outgroup response from the local population. In both cases, the goal of the troll is to draw the outgroup into the antagonistic discussion as if it were a worthwhile and serious controversy. As for the ingroup audience, it can join in the trolling and contribute to it, thus building a coalition of complicity with the original troll. Here, the goal is to entertain the insiders, who know all along that the troll is “merely trolling,” thus reinforcing the ingroup’s sense of belonging and “getting it.” Playing one group off against the other in this way is the basic rationale behind trolling (Barney, 2016; Cohen, 2017; Connolly, 2022; Morgan, 2023).¹⁸

As such, trolling is a paradigmatic instance of many-to-many communication, as it needs different kinds of agents to engage with one another to set off its vicious communicative dynamic. Arguably, there is a precedence to this dynamic in any offline *doublespeak*: inside jokes, dogwhistles, coded messages, etc. (see Herbert and Kukla, 2016). But in the case of online trolling, it is further amplified and sustained by the affordances of many social media and online discussions: anonymity/pseudonymity;¹⁹ self-identification as a troll who, if successful, builds an online reputation as someone obnoxious, while staying anonymous in real life; the sheer reach of the troll, who can possibly engage and deceive thousands of audience members, etc. This makes impactful trolling something rather unique to many-to-many *online* communication.

¹⁷ Connolly distinguishes between the *target* of trolling, “an audience that trolling is performed to,” and the *onlookers*, “an audience (which includes the troll) it is performed for” (Connolly, 2022: 405), whereas Morgan distinguishes between the *victim*, “whomever or whatever is derogated by the troll’s utterance,” and the *target*, “the intended audience of the troll’s utterance,” the latter divided into “non-troll targets” and “fellow troll targets” (Morgan, 2023: 3408).

¹⁸ Curiously enough, Herbert and Kukla (2016) do not consider trolling among the often pernicious ingrouping–outgrouping practices of *peripheral speech* defined as “the informal, typically playful, insider speech that forms the marginalia and the glue holding together the web of discursive practices of a community” (2016: 577). However, they exemplify peripheral speech with “inside jokes, various kinds of slang, some sorts of gossip, informal insider memes and stylistic flourishes, sometimes slurs that have been repurposed for within-community use” (2016: 582) and stress that “to take peripheral speech seriously is to give it an uptake that marks you as an outsider; insiders are those who “get it” as they show by playing along skillfully” (2016: 594). We take this analysis of peripheral speech as coming very close to grasping the general pragmatic structure of trolling.

¹⁹ Morgan (2023) identifies the contrast between the troll’s “commitment to anonymity” and the “target’s lack of anonymity” as one of the key features sustaining the dynamic of a form of trolling, *subcultural trolling*.

Recent philosophical literature has attempted to characterise trolling as a speech act. Connolly (2022) has defined it as a *perlocutionary* act based on the abuse of *seriousness*, while Morgan (2023) has defined it as an *illocutionary* act abusing *sincerity*. There are important nuances to these approaches, and they cannot be seen as a unified theory of how trolling works. But one thing they have in common is their focus on the troll's artful management of inconsistent communicative commitments toward the ingroup and the outgroup. We agree that this is indeed crucial but propose, instead, a different account.

For us, trolling is not a speech act. It is an abuse of speech acts, comparable to lying, which, too, is not a standalone speech act but an intentional abuse (in the sense of Austin, 1962) of a speech act of assertion. A liar ostensibly participates in the practice of assertion but unostentatiously violates the condition that a speaker should only assert things they believe to be true (Saul, 2012; Stokke, 2013). Similarly, the troll's abuse lies precisely in the fact that, on the surface presented to the outgroup, the speech act performed is just like any other serious conversational act, typically assertion, but also any directive, commissive, declarative, or expressive speech act (a troll may be stating, exhorting, promising, etc.). Yet, this ostensible illocutionary force presented as seriously performed to the outgroup is accompanied by another goal, that of inviting the ingroup to see through the abuse to their satisfaction, to have fun, and even join in the trolling. As such, it is a particularly pernicious form of illocutionary pluralism (Lewiński, 2021). Indeed, the overarching goal, or at least a consequence of a troll's acts, is to sow division, disagreement, and polarisation between the ingroup and the outgroup.²⁰

Arguably, then, the whole point is precisely neither to *just* deceive the outgroup nor to *just* connive with the ingroup to poke fun at the outgroup, but the very division between the two, between us and them. In this way, while constantly seeking and grabbing attention, trolling contaminates engagement and, eventually, corrupts communicative trust—and does so at a scale only massive online conversations can do. Therefore, patterns and skills of attention–trust–engagement different from regular offline conversations need to be implemented. One of them is the ability to ignore trolls, hailed as an important argumentative *virtue* (Cohen, 2017); another is the ability to actively challenge trolls (Connolly, 2022). Yet again, given the nature of many-to-many communication, such skills and responsibilities are somehow distributed: Who bears the responsibility for ignoring or challenging trolls—the victims, the outgroup, or the ingroup—is a philosophically important question that only makes sense against the background of the many-to-many conversational model.

²⁰ While we don't take trolling to be a proper speech act with a unique, identifiable *illocutionary* point, with some tweaking Connolly's (2022) approach focused on the *perlocutionary* effects of "the speech act of trolling" is not inconsistent with the pluralistic illocutionary account sketched here.

6. Conclusion

One of our main claims in this chapter is that, to properly understand online conversations, we need something like the many-to-many model of communication. Online conversations display complex patterns of communicative engagement that cannot be suitably analysed in simple one-to-one and one-to-many models. From this perspective, we have argued that online conversations, as they allow for large-scale, many-to-many interaction, have a number of specific features that are less conspicuous in similar offline conversations:

- Participation has the structure of attentional networks.
- People make constant choices of whom to engage with, based on factors such as the perceived trustworthiness of the potential interlocutor and the expected “payoffs” of such conversations.
- Online conversations operate with explicit incentives (in particular, point-scoring and reputation/popularity), which are supported by specific affordances of online environments.
- Common ground cannot be assumed; it needs to be constructed and is more volatile than in one-to-one conversations.

This novel approach to online conversations sheds new light on phenomena such as gamification and trolling, which (or so we argue) can only be suitably approached from a many-to-many perspective that problematises how networks of attention and contrastive attributions of trustworthiness shape conversations. We hope that the conceptual framework introduced here will prove fruitful to study a wide range of other online phenomena, as it is meant to provide a general template for the analysis of online conversational environments.

Moreover, we have argued that many-to-many conversations also occur in *offline* environments, even if they are greatly amplified (both in scale and in regularity) by the affordances of online technologies. Examination of conversations online thus reveals something important about conversations more generally, making them significant objects of philosophical attention, requiring further scholarly engagement. The novel technological affordances of online communication enable many-to-many conversations “on steroids,” thus making them highly conspicuous. But many of the features described here apply to conversations across the board—it is just that it takes online conversations for us to see this more clearly.

Acknowledgments

Marcin Lewiński’s research is supported by the CHIST-ERA/FCT grant “Argumentation-driven explainable artificial intelligence for digital medicine”

(CHIST-ERA/002/2019). Catarina Dutilh Novaes's research is supported by the European Research Council with grant ERC-2017-CoG 771,074 for the project "The Social Epistemology of Argumentation."

References

- Aakhus, M. (2017). The communicative work of organizations in shaping argumentative realities. *Philosophy & Technology*, 30(2), 191–208.
- Aakhus, M., & Jackson, S. (2005). Technology, interaction, and design. In K. Fitch & R. Sanders (Eds), *Handbook of Language and Social Interaction* (pp. 411–436). Mahwah, NJ: Lawrence Erlbaum.
- Aakhus, M., & Lewiński, M. (2011). Argument analysis in large-scale deliberation. In E. Feteris, B. Garssen, & A. F. Snoeck Henkemans (Eds), *Keeping in Touch with Pragmatic Dialectics* (pp. 165–183). Amsterdam: John Benjamins.
- Archer, A., Cawston, A., Matheson, B., & Geuskens, M. (2020). Celebrity, democracy, and epistemic power. *Perspectives on Politics*, 18, 27–42.
- Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press.
- Barney, R. (2016). [Aristotle], On Trolling. *Journal of the American Philosophical Association*, 2, 193–195.
- Beaver, D., & Stanley, J. (2019). Toward a non-ideal philosophy of language. *Graduate Faculty Philosophy Journal*, 39(2), 501–545.
- Beaver, D., & Stanley, J. (2023). *The Politics of Language*. Princeton: Princeton University Press.
- Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F., & Secret, A. (1994). The world-wide web. *Communications of the ACM*, 37(8), 76–82.
- Bianchi, C. (2019). Asymmetrical conversations: Acts of subordination and the authority problem. *Grazer Philosophische Studien*, 96(3), 401–418.
- Blevins, J. L., Lee, J. J., McCabe, E. E., & Edgerton, E. (2019). Tweeting for social justice in# Ferguson: Affective discourse in Twitter hashtags. *New Media & Society*, 21(7), 1636–1653.
- Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- Chen, A. (2015). Unfollow. The New Yorker, November 15. <https://www.newyorker.com/magazine/2015/11/23/conversion-via-twitter-westboro-baptist-church-megan-phelps-roper>.
- Clark, H. H. (1992). *Arenas of Language Use*. Chicago, IL: University of Chicago Press.
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Cohen, D. H. (2017). The virtuous troll: Argumentative virtues in the age of (technologically enhanced) argumentative pluralism. *Philosophy & Technology*, 30(2), 179–189.
- Connolly, P. J. (2022). Trolling as speech act. *Journal of Social Philosophy*, 53(3), 404–420.
- Cook, K. S. (2013). Social exchange theory. In J. DeLamater & A. Ward (Eds), *Handbook of Social Psychology* (pp. 6–88). Dordrecht: Springer.
- Cooney, G., Mastroianni, A. M., Abi-Esber, N., & Wood Brooks, A. (2020). The many minds problem: Disclosure in dyadic versus group conversation. *Current Opinion in Psychology*, 31, 22–27.
- Dingemanse, M. et al. (2015). Universal principles in the repair of communication problems. *PLOS ONE*, 10(9), e0136100.

- Dutilh Novaes, C. (2020a). *The Dialogical Roots of Deduction: Historical, Cognitive, and Philosophical Perspectives on Reasoning*. Cambridge: Cambridge University Press.
- Dutilh Novaes, C. (2020b). The role of trust in argumentation. *Informal Logic*, 40(2), 205–236.
- Dutilh Novaes, C. (2023). Can arguments change minds? Proceedings of the Aristotelian Society, CXXIII(2), 173–198.
- Dutilh Novaes, C., & de Ridder, J. (2021). Is fake news old news? In S. Bernecker, A. K. Flowerree, & T. Grundmann (Eds), *The Epistemology of Fake News* (pp. 156–179). Oxford: Oxford University Press.
- Edwards, A. (2013). (How) do participants in online discussion forums create “echo chambers”? The inclusion and exclusion of dissenting voices in an online forum about climate change. *Journal of Argumentation in Context*, 2(1), 127–150.
- Egan, A. (2009). Billboards, bombs and shotgun weddings. *Synthese*, 166(2), 251–279.
- Esau, K., Friess, D., & Eilders, C. (2017). Design matters! An empirical analysis of online deliberation on different news platforms. *Policy & Internet*, 9(3), 321–342.
- Forestal, J. (2022). *Designing for Democracy: How to Build Community in Digital Environments*. Oxford: Oxford University Press.
- Franck, G. (2019). The economy of attention. *Journal of Sociology*, 55(1), 8–19.
- Frankfurt, H. G. (2005). *On Bullshit*. Princeton, NJ: Princeton University Press.
- Frost-Arnold, K. (2021). The epistemic dangers of context collapse online. In J. Lackey (Ed.), *Applied Epistemology* (pp. 437–456). Oxford: Oxford University Press.
- Gershberg, Z., & Illing, S. D. (2022). *The Paradox of Democracy: Free Speech, Open Media, and Perilous Persuasion*. Chicago, IL: University of Chicago Press.
- Goffman, E. (1981). *Forms of Talk*. Oxford: Wiley.
- Goldberg, S. (2013). Anonymous assertions. *Episteme*, 10(2), 135–151.
- Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Hawley, K. (2019). *How to Be Trustworthy*. Oxford: Oxford University Press.
- Herbert, C., & Kukla, R. (2016). Ingrouping, outgrouping, and the pragmatics of peripheral speech. *Journal of the American Philosophical Association*, 2(4), 576–596.
- Herring, S. C. (2019). The coevolution of Computer-Mediated Communication and Computer-Mediated Discourse Analysis. In P. Bou-Franch & P. Garcés-Conejos Blitvich (Eds), *Analyzing Digital Discourse* (pp. 25–67). Cham: Palgrave Macmillan.
- Hindman, M. S. (2018). *The Internet Trap: How the Digital Economy Builds Monopolies and Undermines Democracy*. Princeton, NJ: Princeton University Press.
- Ivani, S., & Dutilh Novaes, C. (2022). Public engagement and argumentation in science. *European Journal for Philosophy of Science*, 12(3), 54.
- Jaidka, K., Zhou, A., Lelkes, Y., Egelhofer, J., & Lecheler, S. (2022). Beyond anonymity: Network affordances, under deindividuation, improve social media discussion quality. *Journal of Computer-Mediated Communication*, 27(1), zmab019.
- Jenkins, H., Ford, S., & Green, J. (2013). *Spreadable Media: Creating Value and Meaning in a Networked Culture*. New York: New York University Press.
- Johnson, C. R. (2019). Investigating illocutionary monism. *Synthese*, 196(3), 1151–1165.
- Johnson, C. R. (2023). Some varieties of illocutionary pluralism. In L. Caponetto and P. Labinaz (Eds), *Sbisà on Speech as Action* (pp. 121–141). Cham: Palgrave Macmillan.
- Khoo, J. (2021). Code words. In J. Khoo & R. Sterken (Eds), *The Routledge Handbook of Social and Political Philosophy of Language* (pp. 147–160). New York: Routledge.
- Koehler, D. (2014). The radical online: Individual radicalization processes and the role of the Internet. *Journal for Deradicalization*, 1, 116–134.

- Langton, R. (2015). How to get a norm from a speech act. *Amherst Lecture in Philosophy*, 10, 1–33.
- Langton, R. (2018a). Blocking as counter-speech. In D. Fogal, D. W. Harris, & M. Moss (Eds), *New Work on Speech Acts* (pp. 144–164). Oxford: Oxford University Press.
- Langton, R. (2018b). The authority of hate speech. In J. Gardner, L. Green, & B. Leiter (Eds), *Oxford Studies in Philosophy of Law*, Vol. 3 (pp. 123–152). Oxford: Oxford University Press.
- Levine, T. (2019). *Duped: Truth-Default Theory and the Social Science of Lying and Deception*. Tuscaloosa, AL: University of Alabama Press.
- Levinson, S. C. (1988). Putting linguistics on a proper footing: Explorations in Goffman's concepts of participation. In P. Drew & A. Wootton (Eds), *Erving Goffman: Exploring the Interaction Order* (pp. 161–227). Cambridge, MA: Polity Press.
- Levinson, S. C. (2006). On the human "Interaction Engine." In N. J. Enfield & S. C. Levinson (Eds), *Roots of Human Sociality: Culture, Cognition and Interaction* (pp. 39–69). London: Routledge.
- Levinson, S. C. (2013). Action formation and ascription. In T. Stivers & J. Sidnell (Eds), *Handbook of Conversation Analysis* (pp. 103–130). Malden, MA: Wiley-Blackwell.
- Lewiński, M. (2010). Collective argumentative criticism in informal online discussion forums. *Argumentation and Advocacy*, 47(2), 86–105.
- Lewiński, M. (2013). Debating multiple positions in multi-party online deliberation: Sides, positions, and cases. *Journal of Argumentation in Context*, 2(1), 151–177.
- Lewiński, M. (2021). Illocutionary pluralism. *Synthese*, 199(3–4), 6687–6714.
- Lewiński, M., & Aakhus, M. (2023). *Argumentation in Complex Communication: Managing Disagreement in a Polylogue*. Cambridge: Cambridge University Press.
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, 8(1), 339–359.
- Marcoccia, M. (2004). On-line polylogues: Conversation structure and participation framework in internet newsgroups. *Journal of Pragmatics*, 36(1), 115–145.
- Marsili, N. (2021). Retweeting: Its linguistic and epistemic value. *Synthese*, 198, 10457–10483.
- McDonald, L. (2021). Please like this paper. *Philosophy*, 96(3), 335–358.
- McDonald, L. (2025). Context Collapse Online. In P. J. Connolly, S. C. Goldberg & J. M. Saul (Eds), *Conversations Online: Explorations in Philosophy of Language* (pp. 83–105). Oxford: Oxford University Press.
- McGonigal, J. (2011). *Reality Is Broken: Why Games Make Us Better and How They Can Change the World*. London: Penguin.
- McGowan, M. K. (2019). *Just Words: On Speech and Hidden Harm*. Oxford: Oxford University Press.
- McGowan, M. K. (2023). On media reports, politicians, indirection, and duplicity. *Topoi*, 42(2), 407–417.
- McQuail, D., & Windahl, S. (1993). *Communication Models for the Study of Mass Communications*. New York: Routledge.
- Meredith, J. (2017). Analysing technological affordances of online interactions using conversation analysis. *Journal of Pragmatics*, 115, 42–55.
- Morgan, A. (2023). When doublespeak goes viral: A speech act analysis of internet trolling. *Erkenntnis*, 88, 3397–3417.
- Nguyen, C. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141–161.

- Nguyen, C. (2021). How Twitter gamifies communication. In J. Lackey (Ed.), *Applied Epistemology* (pp. 410–436). Oxford: Oxford University Press.
- O'Connor, C., & Weatherall, J. (2019). *The Misinformation Age*. New Haven, CT: Yale University Press.
- Paglieri, F. (2017). A plea for ecological argument technologies. *Philosophy & Technology*, 30(2), 209–238.
- Pomerantsev, P. (2019). *This Is Not Propaganda*. London: Faber & Faber.
- Roloff, M. (2015). Social Exchange Theories. In C. R. Berger, & M. E. Roloff (eds) *International Encyclopedia of Interpersonal Communication*. Hoboken, NJ: Wiley.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735.
- Saul, J. M. (2012). *Lying, Misleading, and What Is Said: An Exploration in Philosophy of Language and in Ethics*. Oxford: Oxford University Press.
- Saul, J. M. (2024). *Dogwhistles and Figleaves: How Manipulative Language Spreads Racism and Falsehood*. Oxford: Oxford University Press.
- Sbisà, M. (2013). Some remarks about speech act pluralism. In A. Capone, F. Lo Piparo, & M. Carapezza (Eds), *Perspectives on Pragmatics and Philosophy* (pp. 227–244). Cham: Springer.
- Schmidt, A. L., Zollo, F., Scala, A., Betsch, C., & Quattrociocchi, W. (2018). Polarization of the vaccination debate on Facebook. *Vaccine*, 36(25), 3606–3612.
- Smith, L., & Archer, A. (2020). Epistemic injustice and the attention economy. *Ethical Theory and Moral Practice*, 23(5), 777–795.
- Stalnaker, R. (2014). *Context*. Oxford: Oxford University Press.
- Stanley, J. (2015). *How Propaganda Works*. Princeton, NJ: Princeton University Press.
- Starr, P. (2020). The flooded zone: How we became more vulnerable to disinformation in the digital era. In W. L. Bennett & S. Livingston (Eds), *The Disinformation Age*, 1st edn (pp. 67–92). Cambridge: Cambridge University Press.
- Stokke, A. (2013). Lying and asserting. *Journal of Philosophy*, 110, 33–60.
- Sun, N., Rau, P., & Ma, L. (2014). Understanding lurkers in online communities: A literature review. *Computers in Human Behavior*, 38, 110–117.
- Sunstein, C. R. (2007). *Republic 2.0*. Princeton, NJ: Princeton University Press.
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.
- Williams, D. (2022). Identity-defining beliefs on social media, *Philosophical Topics*, 50(2), 41–64.
- Witek, M. (2021). Illocution and accommodation in the functioning of presumptions. *Synthese*, 198(7), 6207–6244.
- Wittgenstein, L. (2001). *Philosophical Investigations*. 3rd ed. G.E.M. Anscombe & R. Rhees (Eds.), G.E.M. Anscombe (trans.). Oxford: Blackwell. (Original work published 1953.)
- Yoon, S., & Brown-Schmidt, S. (2019). Audience design in multiparty conversation. *Cognitive Science*, 43, e12774.
- Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. *Social Semiotics*, 25(3), 274–291.
- Zollman, K. (2013). Network epistemology: Communication in epistemic communities. *Philosophy Compass*, 8, 15–27.