The Promise and Pitfalls of Online 'Conversations'

SANFORD C. GOLDBERG

Abstract
Good conversations are one of the great joys of life. Online (social media) 'conversations' rarely seem to make the grade. In this paper I use some tools from philosophy in an attempt to illuminate what might be going wrong.

1. Introduction

Many of our conversations these days take place on social networks: Facebook, Twitter, …. In this respect, these networks are incredibly powerful tools: they provide a virtual 'place' for people to share things of interest to them with others all over the world in real time – whether the shared items are their own (verbally-expressed) thoughts, articles they found interesting, photographs, videos, or what have you. The social networks hold out the prospect of diminishing the significance of geographical constraints and democratizing the ability to get one's word out. And yet many people these days lament the poor quality of our exchanges in online networks.

In this paper I want to use a model of conversation from philosophy of language to diagnose some of the things that seem to be going wrong in our online 'conversations'. The model that I will introduce – what we might call the Stalnaker-Roberts' model of conversation[1] – is a rather simple model, and it employs some very idealized assumptions about the aim of conversations and the conditions under which they take place. As idealized and simple as it is, though, I think this model enables us to shed light on some of the challenges of online discourse, and it brings a fresh perspective on these. In fact, it might be *because* of the idealizations that the model sheds light on the challenges of our online conversations: many of the assumptions that the model needs to make either fail to hold in online conversations or it will be decidedly unclear to the conversational participants themselves whether these conditions hold. The fundamental claim I will

[1] See Stalnaker (1978; 2014) and Roberts (2012; 2018).

doi:10.1017/S1358246121000023

be making is that in this way we can shed light on at least *some* of the problems characteristic of our online discussions.

2. A conversational model

The Stalnaker-Roberts model of a conversation consists in (i) a model of the context of the conversation, (ii) the set of moves that can be made in a conversation, characterized in terms of their standard effects on context, and (iii) the participants in the conversation. We can get a sense of what this amounts to by oversimplifying a bit. Starting with (i), the context, this is understood to consist of (a) the set of propositions mutually accepted by all parties, (b) the question(s) under discussion, and (c) the plans in place, if any. Corresponding to this, the set of moves available to participants, (ii), is taken to involve three main types of acts: (a*) making an assertion, in which one proposes to add the propositional content of the act to the set of mutually accepted propositions; (b*) asking a question, in which one proposes to add a question to the set of questions under discussion, whether as a new question to be addressed, a subquestion of a previous question, or a clarification of an aspect of previous discourse; and (c*) issuing an instruction, in which one proposes to add to the list of plans (either immediate, or more long-term). In each case, those who observe one of these acts face the choice whether to allow the act to have its aimed-at effect on the context – that is, whether to accept the assertion (and so add its content to the set of mutually accepted propositions), and so on with questions and instructions as well.

This (highly oversimplified version of the) Stalnaker-Roberts model of assertion is helpful for thinking about the role of conversation in inquiry. Inquiry starts when there is a question to be addressed before a group of individuals, it involves giving and receiving instructions about how to plan the inquiry together, it proceeds as people add more information to the stock of information that is accepted, and it terminates when all but one of the answers to the main question under discussion is ruled out (leaving that answer as the response to the question).

Still, the model is rather simple, and it makes various simplifying assumptions. I what follows I will bring out some of these assumptions by trying to apply this model to the case of online 'conversations'; in some cases the inability to apply the model to these conversations will point to a shortcoming of the model itself, but in others it will illuminate features of online exchanges that make them less productive, and often less enjoyable, than we might like.

3. Three challenges

I want to highlight three challenges we face in applying this sort of model to online exchanges: (1) it is unclear how often the aim of online exchanges is correctly characterized as a conversation in the sense characterized by this model; (2) it is unclear how to characterize the scope of the participants; and (3) it is unclear how well the acts identified in the model correspond to the scope of acts performed online. In each of these ways, it is hard to see how to apply the model to online exchanges. After arguing that this may well point to sources of unproductiveness in online exchanges, I go on in the next section to argue that these sources of unproductiveness combine with certain aspects of the epistemological dimension of these exchanges, resulting in still further problematic features.

3.1 The Aim of Online Exchanges

Does the aim of online exchanges amount to or involve the sincere exchange of information? Though it might seem so at least for a good deal of our engagements online, a number of researchers have called into question how central an aim this is. Instead, researchers have pointed out how a good deal of online engagement has an expressive aim (Lynch, 2019), or alternatively aims at establishing one's group affiliation. If true, this (by now familiar) point has far-reaching implications: whereas standard conversations aim at or centrally involve the sincere exchange of information with the goal of arriving at the truth, online 'conversations' whose *raison d'etre* is establishing one's group affiliation need have no interesting connection to the truth. What is more, if the aim is to reinforce group cohesiveness, it would seem that participants will restrict themselves to what serves that aim, with the result that contributions that are seen to solidify one's identity will tend to get expressions of approval from others (likes and the like). This will loom large in the next section, when I consider the epistemic dimensions of online exchanges.

3.2 Scope of Participants

Consider ordinary face-to-face conversations. Typically, the participants know who is in the conversation and who is not. In part this is because there are conventions that enable us to discern the contours of the conversational participants. Thus, there are conventional ways

https://doi.org/10.1017/S1358246121000023 Published online by Cambridge University Press

to initiate a conversation with another (or others): for example, by addressing oneself to them in a way indicating one wishes to initiate a conversation with them. There are conventional ways to signal that one is or remains a participant in an ongoing conversation: for example, by shaking one's head to indicate uptake (or performing one or another action indicating that one is attending to the contributions). And there are conventional ways to conclude a conversation. What is more, in face-to-face settings, one can see all of those who are potentially participants in the conversation, enabling one to get a good deal of information regarding the participant status of each candidate.

The salient contrasts with online exchanges are many. For one thing, unless an individual contributes to the exchange (by posting, reposting, replying, liking the post, or what-have-you), there is no way to tell whether an individual in one's social network is a silent 'part' of the conversation. The contours of the participants, then, are hard to discern even if we assume that they are determinate. But there are grounds for doubting whether the contours are determinate: is one who glances at a post or a thread, registering a comment or two only to move on quickly, a part of the conversation or not? We might stipulate that only those are part of the online conversation who explicitly contribute to it in one way or another. But in that case our stipulation would eliminate the analogue in face-to-face encounters of the silent participant: the one who tracks what is going on, updates the context accordingly, but does not otherwise contribute.

Consider, too, the conventions available through which to signal an interest in initiating or concluding a conversation, as well as those for signaling one is still a participant. To be sure, posting might be considered a way to signal an interest in initiating a conversation: but with whom? With everyone in one's social network? With those on whose wall one's post is seen? We might also think of tagging as a way of calling one's attention to the conversation in the hope of including them; but not all tags are for that purpose, of course, so this signal is noisy. When it comes to indicating one's status as a participant in the conversation, of course, one can contribute to it (commenting, reposting, liking, etc.). Still, as signals of participation these acts (of commenting, reposting, liking, and the like) are also noisy, and it remains true that there would seem to be no way of doing so silently. Finally, there is no natural or conventional way to signal the end of the conversation, and it often happens that old posts (left for finished by those participating at the time) are revived when a new contributor makes a very belated contribution.

All of these differences are exacerbated by the temporal dimension of online exchanges: while technology would permit one to see who

among the active participants is online at any given time, in a good many exchanges one interacts only with those who are actively participating at the time. Participants then can be remote from one another not only in space but also in time: I might reply tomorrow to a comment you made today, keeping alive the conversation even though the intervening 24 hours have been non-active.

There is an all-important corollary of not being able to discern the set of participants in an online discussion: we cannot discern the context set itself, that is, the set of propositions that are being mutually assumed for the sake of the exchange itself. The nature of the difficulty here can be characterized by contrasting two extreme views as to how to go about addressing this: what I will call a minimalist and a maximalist approach. On the *minimalist* approach, one assumes the minimal number of participants consistent with the nature of the exchange itself. These will include all and only those who have responded to another's post (whether by commenting or liking or reposting or what-have-you). Here, what is presupposed is only what we need to treat all of *these* individuals as mutually assuming in order to make sense of the exchange itself. On the *maximalist* approach, one regards the speaker's *entire social network* as in on the exchange, and what we as theorists regard as presupposed is what we need to make sense of the possibility of *anyone* of these individuals participating in this exchange. Obviously, how much this includes will depend on the diversity of views within one's social network, the salience of those views, and so forth.

I submit that neither the minimalist nor the maximalist approach to context-fixing is the right way to capture what is going on in the online "conversation," and that the reason for this will generalize to other (less extreme) attempts to capture the set of things that are mutually presupposed in the exchange. The trouble is that neither of these options appears to set the right constraints on what is to count as mutually presupposed in context, and so both will err in including either too much or not enough in the context set itself. Generalizing from this, I suspect that any attempt at some hard-and-fast technique for discerning the context set will make errors of one of these two sorts – either including too much in the context set, or not enough – even if the approach itself is not as extreme as either the minimalist or the maximalist approaches just described.

Let me give examples from the US and the UK to illustrate these worries about the minimalist and the maximalist approach to context-fixing.

Start with the minimalist approach. Suppose you reside in the US and you are among the majority of US voters who disapprove of

President Trump (56% as of July 2020[2]), where the majority of your network disapproves of him as well. Still, you might have a few pro-Trump people among your network, and they might be outspoken on those occasions when you make anti-Trump comments. If they do, does their participation in the thread you've initiated inviolate your attempt to presuppose (for the purpose of discussion) such things as that Trump is aiming to normalize political practices and behaviors that ought not to be normalized? Well, the cost of allowing this would be to allow all of the trolls to set the terms of our discussion. But if we insist that this should not be allowed, then it follows that the mere fact that a person is participating, or is trying to participate (e.g. by making comments on one's discussion thread), is not sufficient, by itself, to include their perspective as serving to fix the context set. To be sure, we might try to rectify this problem by de-friending trolls. But the problem is deeper than that: a troll who makes as if to participate properly throughout doesn't count as preventing one from presupposing what one wants to presuppose merely because the troll is quiet about matters. It seems that the mere fact that one is participating in an online discussion doesn't yet determine the role one plays in fixing the context set. Since minimalism assumes otherwise, it is not adequate.

Move to the maximalist approach. Suppose you reside in pre-Brexit UK and are a firm Remainer, where the vast majority of your online social network is in favor of Remain. (The point doesn't depend on the details of the politics; I use them for the purpose of illustration only.) Still, you might have the occasional Brexiteer among your network. (Suppose they are usually quiet and don't participate much, if at all, in your discussions.) Do their pro-Brexit views nevertheless help to fix what is mutually taken as presupposed for the purpose of your discussions? If so, you will not be able to have a discussion in which the Remain position is mutually taken for granted. But this seems weird: even if you don't always expect to be able to take that for granted, surely *sometimes* – in some online exchanges you initiate – you want to be able to do so, and you expect to be able to do so. So it can't be that the mere fact that you have a few Remainers among your social network prevents you from ever being able to do so. The maximalist approach can't be right.

What I think actually happens: we construct the context on the fly. Some posts make clear the sort of audience they have in mind: one's professional colleagues, or family members, or high school friends, or

[2] This statistic is taken from the web site FiveThiryEight.com, cited 9 July, 2020.

https://doi.org/10.1017/S1358246121000023 Published online by Cambridge University Press

the politically like-minded, or fellow cat-lovers, or fellow members of an interest group of some kind, etc. Other posts are sufficiently general that they might be aimed at a much wider audience, where the contours of that audience are themselves not clearly conceived in advance. Over the course of the evolution of the discussion, participants construct the context set as needed to make sense of the exchange. Insofar as some are regarded as calling into question the intended presuppositions of the conversation, they are ignored or disallowed to continue to 'hijack' the discussion. But if this is correct, it makes clear that there will be many cases in which the state of the context set at a given time will be far from clear to the participants, even to those centrally invested in a productive discussion.

3.3 Scope of acts performed online

To introduce the problems surrounding the scope of the speech acts performed online, I will need first to present some basic elements of speech act theory. To begin, note that the verbal use (or utterance) of a sentence is not just the production of sounds; it is rather a meaningful use of speech. Thus if I utter 'You will sit next to me' to you, intending thereby to be expressing what that English sentence says, I have performed a meaningful act. We can designate this act – the meaningful act one performs when one produces a sentence intending to be expressing what the sentence says – as a *locutionary* act. In knowing which locutionary act I have performed, you (my intended audience) thereby know what I have said. Still, as my intended audience you can know what locutionary act I have performed, and so know what I have said, without knowing how to *take* or *understand* what I've said: I might have said this as a *prediction* (I am predicting you won't know anyone else at the party), a *decision* (we are making seating arrangements for the upcoming wedding), or a *command* (I say it to you under my breath in a threatening tone). We can use 'illocutionary force' to designate that feature of a speech act that pertains to *how* what is said is to be taken or understood by the audience. Thus the illocutionary force of a prediction (= the way the speaker intends the audience to understand her locutionary act) differs from that of a decision, which in turn differs from that of a command. And we can use 'illocutionary act' to designate the resulting type of acts themselves: predictions are a different type of illocutionary act than are decisions or commands. The case above makes clear that one and the same (type of) locutionary act might be associated with various distinct types of illocutionary act.

For its part, the Stalnaker-Roberts model of conversation allows for three general types of illocutionary act: *assertions* (proposals to add information to the stock of propositions that re mutually presupposed), *questions* (proposals to add a query to the stock of questions under discussion, including subquestions of questions currently on the list as well as clarificatory questions about previous moves or other questions), and *directives* (proposals to add an action on the list of what is to be done). Using the notions introduced above, we can say that assertions constitute a type of illocutionary act with its own distinctive illocutionary force (= *assertoric* force), questions constitute a type of illocutionary act with its own distinctive illocutionary force (= *interrogative* force), and directives constitute a type of illocutionary act with its own distinctive illocutionary force (= *directive* force). The challenge is that there are actions performed in online settings whose illocutionary force is not obviously any one of these; and in addition even when it is clear (more or less) that an online act is of one of these three illocutionary types, it appears to be significantly different than standard acts of that type (in face-to-face settings). I will take these up in order.

There are many acts performed in the context of online discussions of which it is not obvious that their illocutionary force is one of the three just described. Here I mention four: posting, reposting/retweeting, liking, and (hash)tagging.

Of all of the 'speech acts' performed online, the post is the one that might seem the easiest to incorporate into the Stalnaker-Roberts model: isn't the act of posting simply the act of assertion itself – at least when one's post purports to say how things are? (By this I mean to exclude posts that are clearly intended as venting, or as merely expressive in some other way, as well as posts that extend invitations, etc.). There is much to recommend this analysis regarding posts in which one purports to say how things are. Still, there are two complications that are worth highlighting, as both of these render the construal of such posts as assertions less than fully happy.

Consider the question that one finds next to one's name when one signs on to one's Facebook account: *What's on your mind, [name]?* Interestingly, this question permits of at least two distinct readings: what I will call the *expressive* reading, and what I will call the *topical* reading. According to the expressive reading, the question asks one to *give expression to* one's own state of mind – whether that involves something one is thinking about, or an emotion one is feeling, or a reaction one is having, etc. According to the topical reading, the question asks one to *address oneself to a topic* and say something about that topic.

https://doi.org/10.1017/S1358246121000023 Published online by Cambridge University Press

Suppose that a Facebook user interprets this question in terms of the expressive reading, and that her posts are informed by this aim. Then she might take it that what she is doing is simply giving expression to her states of mind. To be sure, we can still see her as making assertions in this case. Only if this is how she intended her post, its content will pertain in the first instance to her state of mind, rather than to the topic she is addressing. To illustrate, suppose she posts 'My friend Tom has misbehaved', intending thereby to be expressing 'what's on her mind' (as Facebook would put it). Then she will intend for her post to be understood as capturing e.g. the irritability she is presently feeling in response to something she takes Tom to have done. Even if it makes sense to regard her as having *asserted* as much (namely, that she is irritated etc.), such an assertion is very different from an assertion that is straightforwardly about Tom's behavior. To see this, notice that she might be taken aback by anyone who questions her: she regards herself as having done nothing more than having expressed her own state of mind, including her own take on the world, and any attempt to question this would, in her mind, be seen as challenging her authority to say what is on her mind. This is a very different sort of activity than the one we engage in when we make assertions about the world. To be sure, her 'take' on the world might be called into question (as in: Tom didn't do what she took him to have done); but if she is pressed with such an objection, she can always resort to the response, 'Well, this is how things struck me, and that's all I was saying in my post'.

I mention this not to defend this sort of maneuver, but rather to point to the possibility of some unclarity as to what, precisely, one is doing when one posts on Facebook. My claim is that this unclarity remains even if we restrict ourselves to posts in which one purports to say how things are. And so, even after we agree that posts are assertions in the Stalnaker-Roberts' sense, this possible unclarity makes it unclear what it is that is being asserted in any given case.

There is one other aspect of posting that makes it somewhat hard to accommodate it within the Stalnaker-Roberts model. Whereas that model aims to capture face-to-face exchanges between conversational participants, posts can have the feel of *public announcements* rather than *contributions to a conversation* itself. That is, one who posts is doing something more like *broadcasting to a wide (indeterminate) audience*, than *talking to a determinate set of individuals*. If this is right, of course, then the whole Stalnaker-Roberts model is not applicable in the first place – but then again, neither would it be correct to say that we engage in conversation online. I do not raise this to endorse the idea that we do not have conversations online.

On the contrary, I do think we engage in (something very much like) conversations online; my present point is only that they have very different features than those of face-to-face conversations. Seen in this light, posts can have the feel like public announcements more than they are claims made in the give-and-take of a speech exchange is one such feature.

But posts are not the only 'speech acts' online that are hard to fit into the Stalnaker-Roberts model. Consider next the *repost* or *retweet*, when another person's post, as such, is 'forwarded' in one's name. These are often (typically?) interpreted as a re-affirmation the content of the original post or tweet. And if things were this simple, reposts/retweets could be seen as a kind of assertion in which one re-asserts a content previously asserted by another – presumably with the point of extending the dissemination of that information to one's own social network. The difficulty is that not all retweets *are* endorsements. There are many motives for retweeting or reposting a previously-made post. At its most generic level, the rationale is that of bringing something to the attention of one's social network; but one can have all sorts of reasons for wanting to do this, not all of which include endorsement. (Perhaps it will be obvious to the most salient members of one's online social network that one regards the original post with contempt, or irony; perhaps the point is some sort of 'in'-joke among one's online social network; and so forth.) And even when one does endorse the content, sometimes the point of reposting/retweeting is not to re-assert what was asserted previously, but simply to register or signal *one's own endorsement* of it. (Such an act puts the focus on one's own attitude toward the content posted, rather than on the alleged truth of that post.) No doubt, the difficulty in interpreting a retweet or repost is related to the difficulty of discerning the contours of the conversation (who is included, and who is not), as well as the corresponding difficulty of discerning what is being presupposed in the context. But even if the difficulty of discerning the context is more fundamental, still, the challenge of assigning an illocutionary force to a retweet/repost, and even determining whether that force is one of the three main types postulated by the Stalnaker-Roberts model, remains.

Next, consider the act of 'liking' another's post. This act is even harder to interpret than is the act of retweeting or reposting. What we might call its 'pragmatic significance' – what it intends to convey regarding how it is to be taken by others in the conversation – can be any of the following: I endorse what you've posted; I like what you've posted; I support you in having posted this; I like you; I have read what you posted with interest; you are on my mind as

you post this; I am following this thread with interest; and many others besides. To be sure, the set of possibilities here may be narrowed e.g. on Facebook, where one has other options: in addition to an emoji for 'like', there are emojis for 'love', 'care', 'ha ha', 'wow', 'sad', and 'angry' (the terms are from Facebook). And when the context makes things clear (but see below), interpretations may be narrowed further. Even so, there are a great many occasions on which it is hard to see how to interpret a mere 'like', and in any case it is far from clear how to fit this action into the tripartite list of actions postulated by the Stalnaker-Roberts model. These acts seem more expressive than any of the three acts postulated by that model: far from proposing to add information, or a question, or a task-to-be-done, they merely express one's attitude. No doubt, in this way they can be seen as adding the information that one expressed such an attitude to the common ground; but they are not thereby to be represented as an assertion. (As Stalnaker himself noted, all sorts of things add information to the common ground without counting as assertions: any salient public act one performs will do just this, as will salient events not involving any agent.)

Finally, consider the act of tagging. Here I have in mind tagging on both Facebook (an act involving the use of the person's Facebook name) or on Twitter (an act involving the use of the '@' sign followed by their Twitter handle). (Note that both of these are distinct from the use of the *hash*tag used on Twitter.) The act of tagging someone on a post can be performed with any of a variety of distinct intentions in doing so: to get the target's attention; to indicate to the target that s/he is being discussed; to indicate to the target that s/he is being thought about in connection with the post; to elicit from the target some response; and so forth. Assuming that the relevant intention is discerned by the audience (including but not limited to the target), once again, we might think that the statement that the speaker has the relevant intention is added to the common ground; but again this does not make the act one of assertion. The act would seem more like that of addressing oneself to someone than it would an assertion, though on occasion, when intended to elicit a response from the target, it might be construed as an instruction (to the target to respond). But since this will not cover all cases of tagging, the act of tagging itself is not of a type that should be identified with any of the three types of act postulated by the Stalnaker-Roberts model. (This is not particularly surprising; the point of the act of tagging is not to add any information or question or instruction to the common ground, so much as to capture another's attention for some purpose or other.)

While I will not have much to say about it, the act of hashtagging on Twitter is distinct from the act of tagging (whether on Facebook or Twitter). Though it can be directed at a single individual (whether explicitly, as in hashtagging them by name, or in some other way), the use of the hashtag is typically not so aimed. Instead, hashtagging is aimed at attracting the attention of the widest audience possible. Still, the intentions behind a hashtag can be varied. By their nature, hashtags allow followers of a topic to follow that topic, and might be seen as having that overarching aim. Still one might have various motives (with a greater or lesser degree of openness) for doing so: eliciting a response, capturing attention, adding to a conversation, and so forth. Once again, it is unsurprising that this act is not among the three postulated by the Stalnaker-Roberts model.

4. Disappointing conversations

I have been spending some time trying to highlight some of the features of online 'conversations' that appear difficult to understand in the terms provided by our best account of face-to-face conversations. My aim in doing so has been to prepare the way for an evaluative claim: these features of online exchanges – those highlighted in the previous section – are partially responsible for some of the disappointing outcomes our conversations online. To establish the latter claim, I need to supplement these features with some claims pertaining to the epistemological dimensions of our speech exchanges (whether face-to-face or online). It turns out that, once we understand some of these dimensions, we would predict some of the difficulties and problems that arise in online exchanges. Or so I will be arguing in this section.

A good proportion of the problems that arise in online exchanges reflect our uncertainty as to the context of the exchange itself, where 'context' is understood in terms of the Stalnaker-Roberts model. Such uncertainty, or more generally the failure to track the context as it evolves dynamically throughout the exchange, has at least some explanatory role in such phenomena as (i) the speed, extent, and ferocity of online shaming, (ii) groupthink, and (iii) belief polarization. I want to begin, then, by characterizing the source and nature of our uncertainty regarding the context, and the difficulties involved in tracking the dynamics of context as it evolves throughout an exchange.

I noted above that online exchanges on social media such as Facebook and Twitter make it practically impossible to know the

https://doi.org/10.1017/S1358246121000023 Published online by Cambridge University Press

contours of a conversation: while we can discern some of the participants in an exchange (namely, those who have actively contributed to it), we are unable to determine those who are silent participants to the exchange, those who are following the exchange albeit without weighing in. The implications of this ignorance are hard to overstate.

To see why this is, it will be helpful to identify the members of the class of inferences that are characteristically drawn from audience reactions to mutually observed contributions to a conversation. Here I highlight those inferences pertaining to the acceptability of a speaker's contribution. Suppose you are participating in a large face-to-face conversation in which a speaker is making claims about a topic on which you know little. The claims seem plausible, though you don't know enough even to be confident of your sense of their plausibility. Still, you see others nodding in agreement, and you take this to indicate their sense of the acceptability of what is being said, and you regard this as still further evidence of the likely truth of the claims. On this basis, you accept the speaker's say-so. Here, you are using the audience's apparent agreement as evidence (of a higher-order sort) indicating the acceptability of the speaker's say-so. Alternatively, if you observe that the audience is perplexed, or seems dubious, or is raising doubts, you might take this as some evidence that the speaker's say-so is *not* to be relied upon. In either case you are treating the audience's manifest reaction as offering evidence bearing on the acceptability of the assertion made in your mutual presence. Notice that you might do so even when the audience's reaction is one of silence itself. Reasoning that if they had harbored doubts they might have indicated this, you might think that their silence is attesting to their having accepted the say-so, in which case you are treating their silence as evidence of the acceptability of the say-so.

These are familiar features of face-to-face conversations involving multiple people. But now when these features are combined with our ignorance of who is following an online discussion, we can run into some serious problems. In particular, I suspect that this is partly responsible for such things as (i) the speed, extent, and ferocity of online shaming, (ii) groupthink, and (iii) belief polarization. Let me explain.

Suppose you observe a post on Facebook in which a Facebook friend writes of a situation that she finds worthy of contempt. You are aware that it will be regarded as such by your peer group as well, at least some of whom are Facebook friends with the one who posted. While you do not know what their ultimate views are, you worry about a scenario in which they too regard the situation as contemptible while simultaneously condemning those who don't so

https://doi.org/10.1017/S1358246121000023 Published online by Cambridge University Press

regard the situation. Once you see a friend join the speaker in expressing contempt for the situation, you decide that you too must quickly make clear that you find the situation worthy of contempt – lest others who are aware that you are Facebook friends with the writer might mistake your non-response for not caring. So you speak up in condemnation. Of course, many of your Facebook friends reason in the same way. The result is that many rush to join in the expression of contempt. Of course, once others see so many do so, they fear that not doing so might be taken (by those who think they are following this discussion) as a lack of contempt, and so they soon join in as well. What started out as a single person expressing contempt has become a pile-on.

Notice the role that ignorance of context plays in this scenario. There are various significant aspects of this ignorance. In general, you are ignorant of who is part of the conversation; and you are ignorant of the reactions of those who are part of the conversation but who are silent. The former ignorance might lead you to worry that there are far more participants than those who have been actively participating. The latter ignorance might leave you worrying about what those silent others (whoever they are) are thinking: what they think of the speaker's contribution, and also (perhaps more worryingly) what they think of *your* reaction to the speaker's contribution. All of this comes to a head in the form of the concern that others might take you to be following the conversation and might misconstrue your silence; this often leads you to respond as quickly and as vehemently as you did. And of course what goes for you goes for many others as well. What we have, in short, is a perfect storm in which everyone is ready to pounce on any shameful behavior mentioned online.[3]

I suspect that this sort of 'contextual ignorance' is explanatorily relevant not only to the speed, extent, and ferocity of online shaming, but also to groupthink and belief polarization.[4]

Groupthink is the phenomenon in which members who self-identify with a group shape their attitudes so as to bring them into line with those of the other group members, where this is driven by the desire to remain in the group's good standing. Groupthink itself can be seen in all settings, including face-to-face settings: Insofar as one wants to retain good standing in a group and one thinks that

[3] In highlighting the dimension of ignorance and its role in online shaming, I mean to be supplementing the account of online shaming in Ronson (2015).

[4] See e.g. Goldberg (2017b, 2020) for a detailed description.

https://doi.org/10.1017/S1358246121000023 Published online by Cambridge University Press

this requires adopting a certain attitude, one is likely to adopt that attitude. But online settings can exacerbate the situation: insofar as one is ignorant of the contours of an online discussion and one worries about the inferences others might draw of one's silence, one is likely to signal one's attitude by contributing to that effect in the online discussion; and this only encourages other silent participants both to see the attitude as required by the group, and to manifest that they too possess it (by contributing to the discussion).

Belief polarization is the phenomenon whereby a group of likeminded individuals adopt increasingly radical views, or become more confident in their existing views, after discussion with fellow likeminded individuals, even though no new evidence has been introduced in the course of the discussion.[5] The phenomenon itself is seen in face-to-face discussions. But again online settings can exacerbate the problem. Given the dynamic just described, where an increasing number of people feel the pressure to signal that they, too, hold the view in question, this will give everyone more reason to think that the view is widely shared. And insofar as there is evidence to think that the view is widely shared, this gives those on the fence some reason to question their ambivalence, and it gives those who already have the attitude a reason for further confidence (on the assumption that so many people can't be wrong[6]).

I have just highlighted how the features of online discussions (as outlined in section 3) give rise to a kind of ignorance of the conversational context, with the result that discussions online are often sorry affairs in which (group-enhancing) ignorance proliferates. Stepping back from this, I would diagnose a more general challenge we face in our discussions online: not only is our route to information highly dependent on the say-so of others, but what is more the mechanisms in place to correct that say-so are themselves highly dependent on the say-so of our social network.[7] The result is that these correction mechanisms are only as good as the members of our social network are both knowledgeable and outspoken. If you

[5] That is: no new evidence *beyond the evidence pertaining to what others think on the matter*. For further discussion of the epistemological dimension of polarization, see Goldberg (2017b).

[6] It should go without saying but I will say it anyway: I am not endorsing this reasoning in such cases. My claims are rather that (1) such reasoning is common, and (2) such reasoning does capture something distinctively epistemic, in that evidence of what others think is a kind of evidence after all (even if it can be highly misleading as to what the truth of the matter actually is).

[7] See Rini (2017).

https://doi.org/10.1017/S1358246121000023 Published online by Cambridge University Press

happen to reside in social networks where ignorance rather than knowledgeableness is pervasive, or where those who are knowledgeable do not speak up, you are out of luck. What is more, if you are in such a situation and you rely on the network itself to distinguish who among your network is knowledgeable, you are likely to compound the situation: not only do you fail to have knowledge, what is worse you will be *ignorant of your very ignorance*. If I am right to think that this is the state of many of us these days, it does not make for a happy world.

5. Conclusion

In this short essay, I have used a simplified version of the Stalnaker-Roberts' model of conversational dynamics to illuminate the features of online conversations that might explain why these are often such unproductive and unhappy affairs. I have identified two main sources of such unhappiness. The first source derives from the challenges we face in discerning the nature of contributions we make online in the first place. There is a question in each case whether our contributions are best thought of as contributions to a conversation, as opposed to reactions to a public announcement. And there is the challenge of discerning the (illocutionary force of the) the various acts we perform online (liking, tagging, hashtagging, and so forth). The second source the derives from the profundity of our ignorance of context and its evolution as the "conversation" progresses. I have tried to suggest how these sources combine to make online platforms ripe for the sort of ugly and unproductive exchanges that are all too common in our online exchanges.

Northwestern University
s-goldberg@northwestern.edu

References

Sanford Goldberg, 'Should Have Known', *Synthese*, 194 (2017a), 2863–94.

Sanford Goldberg, 'Can asserting that p improve the speaker's epistemic position (and is that a good thing)?' *Australasian Journal of Philosophy* 95 (2017b), 157–70.

https://doi.org/10.1017/S1358246121000023 Published online by Cambridge University Press

Sanford Goldberg, *Conversational Pressure* (Oxford: Oxford University Press, 2020).
Michael Lynch, *Know-It-All Society: Truth and Arrogance in Political Culture* (New York: Liveright Publishing, 2019).
Regina Rini, 'Fake news and partisan epistemology', *Kennedy Institute of Ethics Journal* 27 (2017), 43–64.
Craige Roberts, 'Speech acts in discourse context'. In Fogal, D., Harris, D., Moss, M., eds. *New Work on Speech Acts*. (Oxford: Oxford University Press, 2018), 317–59.
Craige Roberts, 'Information structure: Towards an integrated formal theory of pragmatics', *Semantics and Pragmatics* 5 (2012), 1–69.
Jon Ronson, *So You've Been Publicly Shamed* (New York: Riverhead Books, 2015).
Robert Stalnaker, *Context* (Oxford: Oxford University Press, 2014).
Robert Stalnaker, 'Assertion' In Cole, P. ed., *Syntax and Semantics*: *Pragmatics* 9 (1978) 315–22.

https://doi.org/10.1017/S1358246121000023 Published online by Cambridge University Press