

Andrei Ionuț Mărășoiu¹

Department of Theoretical Philosophy & Center for
Logic, Philosophy and History of Science
(CELFIS)

University of Bucharest

Sandra Cătălina Brânzaru²

Faculty of Psychology and Educational Sciences &
Research Institute

University of Bucharest (ICUB)

THE PERILS OF DOXASTIC AUTONOMY

Abstract: *Progress in digital technologies, especially in social media use and the internet, offers opportunities and enhances risks for what one knows and understands when compared to everyday communications prior to the age of the internet. Can risks be mitigated and opportunities seized, and to what extent would it be rational to do so? We inquire into the variety of epistemic values the consideration of which might shape the epistemic environment of digital communication so that it may apply both to ideal agents but also to everyday users in non-ideal circumstances.*

Key words: *autonomy; epistemic value; echo chambers; evidence; truthfulness; intellectual character*

¹ andrei.marasoiu@filosofie.unibuc.ro; ORCID: 0000-0001-9400-7702

² sandra-catalina.branzaru@fpse.unibuc.ro; ORCID: 0009-0004-8022-9485

1. Introduction

Digital progress raises questions about our everyday communicative practices. Phenomena such as fake news and the spread of conspiracy theories may be enhanced by digital technology. Echo chambers and communication bubbles, even though they are not the exclusive domain of digital communication, are prominently featured in our online conversations.

This raises a philosophical worry: are we in full control of our own communications? Ideally, we might expostulate that speakers are responsible for what they say just as, by parity of reasoning, agents are responsible for the most salient consequences of their actions, and that parallel is supported by our taking speech acts to be acts *tout court*. The general concern in the background of our text is how to move from an epistemology of such ideally epistemically autonomous agents to an epistemology of social media, brimming with “happy go lucky” users who group and regroup in echo chambers and bubbles – a non-ideal epistemology of ignorance, vice, complacency, banter, gossip, riz, and the like. And the specific question we ask is if all these phenomena might, in non-trivial ways, lead us to reconceive notions of ideal epistemology such as doxastic autonomy.

We will argue that, even though it is among the most prized resources we as a community recognize, there are also perils to doxastic autonomy. Such perils, while not exclusive to digital environments, are in turn enhanced by them, raising twin worries: the practical question of how to mitigate the risks of digital communication; as well as the more reflective concern of how to conceive of

ourselves as epistemic agents in both physical and digital environments.³

Our aim is not primarily to elicit your assent to any unique view of doxastic⁴ autonomy but rather to open up an array of different possible ways to conceptualize dimensions, aspects or strands of doxastic autonomy such that we may gain a slightly more articulate appreciation of what such autonomy comes down to in digital communication.

In appraising doxastic autonomy we intend to remain neutral on a host of foundational questions about it ((Frankfurt 1971), (Buss and Westlund 2018)):

- whether it pertains to acting freely vs. freely intending to act vs. freely forming the intention that then causally prompts action;
- whether it pertains to cognitive or conative epistemic behavior, e.g. coherence-enhancing, integrity-furthering or stability of second-order core preferences;
- how robust the knowledge condition on free will (knowing what one is doing should be understood) such that it captures the seemingly autonomous yet

³ We are not committed to a hard and fast distinction between physical, digital, online and virtual environments. Rather, the epistemic values and their imperfect realizations we explore run the gamut of all such environments. For metaphysical debates surrounding possibly distinct kinds of environments, see Chalmers (2019).

⁴ Fake news and of conspiracy theories often undermine their followers' claims to knowledge; we find it more appropriate to talk about doxastic autonomy rather than epistemic autonomy. The issue is precisely which aspects of doxastic autonomy are preserved when backing into believing a conspiracy theory, in a way that undermined the thinker's general abilities to reap knowledge.

superficial interactions in social media; e.g. we seem to know what we are doing; we are carefree / lack doubt / don't find problematic how we behave online; we might, on reflection, find ourselves to be carefree (lack doubt, etc.) if we were to consider the question but we rarely do consider it; and so on;

Indeed, we will argue that doxastic autonomy might require reassessment in light of our online communicative practices in social media regardless of how one answers these foundational questions concerning the nature of autonomy.

2. Divjak on the risks digital development poses for liberal democracies

It helps to start from Slobodan Divjak insightful perspective on the risks incurred by digital development for liberal democracies in order to raise the question of how digital progress in our communicative technologies might alter, if at all, our doxastic autonomy.

In discussing Divjak on the risks that digital development poses for liberal democracies and authoritarian regimes alike, it seems best to start from a few selected quotes from professor Divjak's own contribution. This is our own selection, and doubtless his contribution is more nuanced than such excerpts could capture briefly. Here are the quotes we selected:

- (1) "It is undeniable that the development of various forms of sophisticated technology enables a far more efficient way of monitoring and controlling the behavior and mutual communication of citizens in their private and semi-public lives" (p. 29)

(2) “In classical totalitarianism, the sovereign dictator, through his secret intelligence services, developed a very wide networks of informers or snitches among the ordinary population, who provided the representatives of these services... with information about ideologically deviant, dissident or oppositional activities and verbal anti-regime statements of their compatriots in their private sphere.” (p.28)

In turn, these excerpts paint a broad picture of both digital technology and communication surveillance in authoritarian regimes. The question is how the two connect. As a possible interpretive sketch, we offer the following argument:

(a) Digital progress enhances information control over epistemically opaque users.

(b) In totalitarianism, information control leads to oppression.

Therefore, by analogy,

(c) Provided state- or corporate-sponsored entities resourceful enough to oppress, information control leads to the oppression of epistemically opaque users in liberal democracies as well.

We are well aware that this argument does not do full justice to professor Divjak’s concerns. That said, it might capture a worry everyday users of social media might have. The restriction to epistemically opaque users ensures autonomy is *in theory* achievable; but we are under no misapprehension that most users are – and will continue to be – ignorant about how their data are exploited in digital communication.⁵

⁵ We appeal to a broad conception of ignorance, multifaceted with respect to technological details, background computer science, transactional details concerning how user data are exploited, etc. We

The key notion at play in both premises is that of information control. We hasten to say that we will not be proposing an explicit technical view of what information control amounts to. That, we believe, depends on the current state-of-the-art and digital communicative technologies; little profit might be drawn from armchair speculation about this. However the technical notions of information and data control (see Section 6 below) both anchor back into the putative autonomy and responsibility users, agents and speakers like ourselves hold for the informational content of what they communicate. It is for this reason that we take the notion of informational control to be key here.

Since this is an argument by analogy, we need to explore what relationship there might be between classical tactics of information and communication surveillance in authoritarian regimes, on the one hand, as well as evasive techniques, nuanced and veiled communications on the part of everyday users seeking to avoid these, and, on the other hand what is, and respectively what ought to be, the behavior of everyday social media users, as well as agents and other forms of digital communication – be they more classical or, on the contrary, open to virtual environments.

3. Echo chambers, fake news and conspiracy theories: working definitions

In discussing the opportunities and risks that digital communication poses, it seems best to focus on social media platforms that are accessible and frequently used by everyday users such as ourselves, with little technical

also stay neutral about what the best way to conceptualize ignorance in such varied cases might be. For discussion, see Alcoff (2007).

or engineering expertise yet desirous to communicate with others – family, friends – who might be far away.

Relevant in what follows are the following half-baked definitions:

- echo chamber: forum on social media where users share similar content among themselves, usually with political implications, filtering users who dissent or whose interventions are dissonant
- fake news: information on social media that is inaccurate, and of which intent to deceive may be reasonably inferred
- conspiracy theory: background story, usually of a fantastic nature, lacking backing by evidence but extremely persuasive for users in the echo chamber wherein that conspiracy theory spreads and is sustained

We hasten to add that these are working definitions, not to be taken as the final word on the matter. They are, rather, signposts that some such characterizations are required in order to be able to then finesse approaches to their target phenomena: echo chambers, fake news and the spread of conspiracy theories.

4. Technology, opacity and undermining autonomy

With these working definitions on board, let's return to the issue of whether progress in digital technologies and communication enhances risks already present in communication prior to the age of the internet – or perhaps, yet more, presents new risks heretofore unknown.

One might think that various forms of addiction (of the internet; of games; news junkies; dependence on online communication for social perception and validation of one's self-worth, etc.) might in and of themselves be harmful. By and large, however, and used with moderation, technology alone usually harms no one. Fake news and conspiracy theories often promote extraneous ends, even if their medium of propagation might be digital or digitally enhanced. However, we should need to consider user vulnerabilities when interacting with such fake news or conspiracy theories. Instead of promoting epistemic values, many of us find ourselves in a position of facing epistemic risk, primarily stemming from epistemic opacity and self-effaced autonomy.

Opacity first. Often, users have no knowledge about the mechanisms of the social media platform they are using – we are not informed about its background information – so there is no way of tracing input on its way to output. There are various reasons why this kind of information is not shared. Either it is too complicated to explain, and users might resist using the app. Or the app allows access to the user data to third party apps – or, in some cases, to the state. The data no longer belongs to the users, even though it is produced by the users.

Users have no information about the algorithmic activity behind the data they are fed on a daily basis. What usually happens is that users end up being exposed to content and information that suits their preexisting beliefs, which, over time, leads to selective exposure, epistemic fragmentation into different epistemic communities – which may be called communication bubbles or echo chambers. Of course, this can be

controlled by epistemically vigilant users who know how the background processes work. Such users often avoid opacity: they control their data to the best of their ability, and most of them don't actually use social media, or if they do, they control the app and the content they share.

Since most users do not know how they get to be exposed to the data they are exposed to (lack of coverability, Goldberg, 2011) they can become subjects of opinion manipulation, since they cannot distinguish between trustworthy verified content, false information, false claims or accidentally generated false claims (if the content is produced by LLMs, LLMs lack the ability to correctly account for sourced they use). This might bear on how the beliefs of such users are formed. So, one interpretation might be that the opacity of algorithmic content generation fed to users might lead to them forming unjustified true beliefs. If they form unjustified true beliefs because of the data they were exposed to, are they responsible for further sharing statements in the digital environment, based on those acquired beliefs? This is one core challenge to their doxastic autonomy, that we will explore in what follows.⁶

⁶ It is worth mentioning that, when exposed to conspiracy theories, some might acquire beliefs that protect them from uncertainty. Such people usually develop false memories around the pre-existing beliefs they held prior to exposure to said conspiracy theory, they forget they were persuaded into believing whatever it is they believe now to be the case (cf. Douglas, Sutton and Cichocka 2017). That said, we intend our concerns about doxastic autonomy to reach more broadly, to reflective and unreflective users alike, whether the specifics of forming false beliefs in self-deception applies to them or not. The grip of risky online communicative doxastic behavior applies to most of us.

5. Data, content and context collapse

Users do not know what apps do with the data, nor who has access to the data (a form of context collapse). Data are interpreted in relation to acts the user has undertaken in the physical world. Data digitally track the user's movement. Social media platforms can allow third parties access to the data. No-one actually controls the data; the user only has control over the content they produce, in most cases. And so the issue of *data autonomy* doesn't arise – or, if it does, most users lack it, typically because they transacted it away, in the agreed upon terms and conditions, to the platforms they use.

It helps to think of this in terms of what each stakeholder may do. Because of epistemic opacity, what most users lack is control over the metadata shared once they share content. Metadata track the user, not the content that the user shares. Third parties, including governments, do not control the data (unless totalitarian, in which case means of communication might be controlled as well). Such actors rather interpret the data and analyze it. The problem arises when they may *use* the data to track the user's deeds.

Social media platform moderators, in turn, only have control over the data. They can control it in the sense that they can do things with it, like share it with others, sell it, give the government access to it. But neither technology, nor admins, nor third-party beneficiaries control information. Control often presupposes some sort of intention to do something, whereas algorithms are not directed to control any kind of information; they just do what they are supposed to, i.e., generate more content based on some criteria.

Platform owners may usually direct algorithms to collect user data (and some users actually share a lot of information), and then to connect the data with algorithms that run ads. The idea is to match the right ads with the user. The same thing happens to all sorts of content. They feed the user with content that the user likes. This leads to lack of coverability, which (in some cases) collapses into echo chambers. This raises the key issue of *content autonomy*, i.e. whether⁷ the content is genuinely or authentically produced or appropriated by the user.

All this applies to the wealth of us, unwitting users. A *general digital autonomy* might be possible for a user who is in the know. But, as we suggested, informed users rarely linger to use platforms the way everyday users do. Such an ideal, while conceivable, is so infrequent as to not make vivid the risks we envisage.

Discussions about privacy and data control have recently been connected with the digital phenomenon of context collapse,⁸ in which multiple audiences, people

⁷ We take no sides in debates concerning authenticity in the virtual environment. Whichever conception of authenticity one prefers might fit the bill, on condition that everyday users recognize themselves (their genuine selves) in its provisions. For discussion, see Moran (2012), Iftode et al. (forthcoming).

⁸ Although the phenomenon of context collapse exists outside of the digital environment as well, we particularly focus on the digital because, as Marwick and Boyd (2011) and Frost Arnold (2021) argue, it comes with different epistemic dangers and benefits. Due to the fact that most people have multimedia gadgets that are connected to the internet, any person who gives a public speech or who has a simple conversation with someone else in a public place, can become the subject of online context collapse if the person is recorded or livestreamed on a social media platform. The audience that the person envisaged while delivering the speech/talk (the people in the

from different contexts in a user's life (work friends, acquaintances, family members or groups unknown to the user) merge into one single audience or group, which becomes the receiver of whatever a user decides to post as a message on a digital social network platform.

What allows for such aggregates is the wide spread of new technologies, the fact that social media profiles are often non-anonymous and the epistemic opacity of social platforms – the control of users over posts is limited, either by the fact that they do not know how to control them or by the fact that no control options are available to them.

The content one posts leaves digital traces linked to past or paralleled online identities available to users' friends, acquaintances, colleagues, relatives; this bears on the power of a user to target specific audiences. Davis and Jurgenson distinguish between context collusion and context collision. The former occurs when individuals "intentionally collapse, blur, and flatten contexts" (Davis & Jurgenson, 2014, p. 480) and the latter (collision) occurs when "contexts come together without any effort on the part of the actor, and sometimes, unbeknownst to the actor with potentially chaotic results" (idem, p. 481). Intentionally updating personal information, such as graduating from a specific program, moving to a new city, would count as context collusion, whereas being

room) can be broader (and unknown to the speaker). This is not necessarily an entirely good or bad thing, it is just that online context collapse can be good or bad in different ways, much faster than its offline counterpart. One of the main challenges is the fact that context collapse can promote misunderstandings, since the background knowledge of the mixed audiences about the speaker or the user who posted the message is different, not shared by its members.

tagged in someone else's posts or photos without notice and approval, or giving a talk at a conference while this is broadcasted on a social platform by a user in the public, would count as context collision.

This phenomenon affects individuals' identities (personal and professional), which can become blurred on social platforms *via* context collapse to the point that potential misunderstandings or reputational risks might occur. Also, narratives presented to an audience can clash with narratives presented to a different audience, in the sense that they no longer appear linear or consistent. In some sense, context collapse can lead to narrative collapse, especially if one follows one-directional persuasion (one tries to present a fact with a story that makes it more graspable for a kind of audience, but no longer commit to the storyline when presenting a fact to a different kind of audience). This will bear on how a fact is framed, understood, and disseminated (facts are decontextualized, making them vulnerable to misrepresentation).

Context collapse takes away from users the control they have over how they perform in front of a specific audience (the one taken to be the actual audience). Someone may choose to share some of their opinions with an audience and to not disclose them to a different audience, for various reasons (e.g., they want to preserve a narrative which might be misunderstood if presented to an audience that lacks the background knowledge of the imagined "actual" audience).

Another challenging consequence of context collapse is that it allows for the formation of unstable frameworks for evaluating truth, which is why it is usually associated with the spread of echo chambers, epistemic bubbles and

is thought to allow for conspiracy theories to be circulated. Statements intended for one audience might not neatly fit into the "truth" or "falsehood" categories when interpreted by another audience, leading to miscommunication or misrepresentation. Some statements may require a specific framework to evaluate their truth, collapsed contexts often strip away these frameworks, making it difficult to evaluate claims accurately. Echo chambers are environments where individuals are exposed primarily to information that reinforces their existing beliefs, while epistemic bubbles occur when individuals are unaware of the information that contradicts their views. Context collapse might disrupt these bubbles by facilitating exposure to a wider variety of opinions. Nonetheless, if the dominant narratives within a context are biased, collapse could reinforce those biases instead by providing an opportunity for users to defend their beliefs.

When it comes to conspiracy theories, a collapsed context can provide visibility and fuel their spread, but it can also lead to increased scrutiny, fact-checking and use of critical thinking in relation to the exposed content. In both cases, the audiences seem to expand, either via a challenging discussion or fact-checking spree, or via spreading the conspiracy theory further. Even the discussion can further occasion polarisation, as some may interpret challenges as evidence of suppression, reinforcing their belief (or the online behaviour according to which they believe) in the theory's "truth". Conspiracy theories often rely on a selective and self-referential system of truth and they create self-referential loops – e.g., if a conspiracy theory posits that all counterarguments are part of the conspiracy, the system

becomes self-reinforcing and hard to falsify. Some elements of conspiracy theories rest on statements that are unverifiable, and they often gain traction by collapsing contexts—taking data or statements from one epistemic realm (e.g., scientific evidence) and reframing them in another (e.g., political or ideological discourse).

The question is (for our contribution here, at least) what resources classical epistemology can marshal in order to address the flexible, often unreflected yet fluent behavior most users of social media evince, especially as it pertains how they come to form and sustain beliefs, share them with others in echo chambers, and dispute amongst themselves with disregard to logical norms but in seemingly systematic ways. What count as doxastically autonomous agents given imperfect communicative media, and how their autonomy shows in their communicative interactions, is of special interest here, not only from the standpoint of media studies or social epistemology, but also pointedly concerning their individual normative profiles in believing.

6. Wittgenstein’s newspaper: from content to evidence and attention

Here is a familiar example nicely captured by Jennifer Lackey:

“The problem with echo chambers is vividly captured by a well-known example from Wittgenstein in which he describes someone assuring himself of the truth of a report by reading several different copies of the same newspaper. Whether you read one or one thousand copies of today’s New York Times, the evidence you take in about today’s news is exactly the same. You can’t increase the epistemic status of your views by simply

consulting multiple versions of the same source.” (Lackey, p. 206)

Wittgenstein’s newspaper is a terrific example for us mortals (maybe not for ideally rational agents; they are optimally informed). For regardless how many copies are printed – abundant data, each copy differs from the rest – the content stays the same. To say that content is picked up by different readers differently is to rehearse the venerable point (Grice 1957) that the literal semantic content expressed by a discourse may often differ from both how it is received and how it was intended. This is key to understanding our everyday activities on social media: post, repost, tweet, retweet, like and share. In some sense relevant here, the same content rolled over platforms, bounced within bubbles until the news cycle throws in a shinier toy.

There, as with Wittgenstein’s newspaper, the problem is clearly not just about newspapers, certainly not just about the New York Times. The same content is tumbled over without concern for its evidential backing. What matters, content being equal, is the appraisal of evidence: e.g. which are verified as authentic and which not. And not in general but about the salience of evidence in echo chambers and in digital communications more broadly.

Here is what we take Lackey’s main point to be: “you can’t increase the epistemic status of your views by simply consulting multiple versions of the same source”. In what follows we’ll controvert this, primarily raising questions about what count as multiple versions and when the source is identical. For instance, is there are several close witnesses to an event, say a crime, presumably the multiple versions provided by each of the witnesses are not immaterial to the case even if the source

is the same, namely what they witnessed. Rather, we would wish to appreciate the variety of witness testimony so as to gauge – by overlaps, internal conflicts and partial accounts – the reliability of each witness, their relative distance to the event they saw and other relevant details. Witnesses are not like newspapers; for even if the content (what they witnessed) might be the same, witness accounts often differ, in contrast to different copies of the same newspaper.

By Lackey's lights, the target notion being explained seems to be the epistemic status of your views, and the *explanans* for this *explanandum*, the evidence you have available in favor of your views and how this balances against countervailing evidence.

It pays off to contrast how Lackey sees the newspaper case with Wittgenstein's own words:

"He may attend to what he reads, or again as we might put it a function as a mere reading- machine: I mean, read aloud and correctly without attending to what he is reading; perhaps with his attention on something quite different (so that he is unable to say what he has been reading if he is asked about it immediately afterwards)" (Wittgenstein, par. 156)

Wittgenstein's focus is elsewhere: You may read while paying attention or not. You may take the evidence in or not. You may require repetition revisiting passages to pay closer attention, rereading to drive the point home, rereading to reflect aloud, and so on. None of this pertains to the evidence there for the taking. All of it, contrariwise, seems to pertain to how you take that evidence in. Generalizing: doxastic autonomy seems to differ from regarding agents as passively taking stock of the evidence whose sources are available. We should do

best to be concerned with how agents pick evidence up (if and when they do), process it, and roll it over (whether they do so critically or naively).

7. Lackey on autonomy in echo chambers

Let us emphasize one key point of agreement with Lackey; she writes: “rarely do we literally behave as a blind follower of a guru, and so there is at least some rudimentary discrimination that goes into our belief-forming, and social-media-sharing, practices. At the very least, in order for us to accept a view, it has to pass through our own belief frameworks, and so the more of these that a view passes through, the more epistemic support it has, even where there is a significant amount of dependence among the beliefs” (Lackey, p. 220).

Here, just as Wittgenstein, Lackey seems acutely aware that agents do play a role in picking up and disseminating information (whether veridical or not) over echo chambers. She rightly conjectures that, ideally, users are not just passive machines that relay input information into output information as an exact replica.⁹ Ideally again, people are (to varying degrees) discerning, evincing points of view with respect to the information that reaches them and that they pass on. And yet, aplenty everyday users do precisely that: some just mindlessly share stuff to kill time, because they are bored, because whatever, they do not need to care or form beliefs about what they share, pass one, or consume.

⁹ To say some machines are passive is not to imply all are. For a discussion relevant to whether current chatbots are any less creative or reflective than everyday users of social media by any species-neutral standard, see Brânzaru (2024). agential propensities, capabilities and acts in response to the evidence available. It is from this standpoint that we explore the contours of epistemic agency in what follows; the issue is not one in epistemology but in metaepistemology.

Even in the ideal case, then, doxastic autonomy is not just a matter of ascertaining sources of evidence for their accuracy and reliability, but rather of properties that can be ascribed to agents/thinkers in light of their communicative performances.¹⁰ Doxastic autonomy seems to pertain to all these dimensions, external *and* internal. In non-ideal cases, doxastic autonomy crucially depends on their in-context balance. (What are the last three pieces of news you read today? What are the last three comments you posted? Try to remember each as accurately as possible, and then ponder our fragile doxastic condition.)

As Lackey aptly puts it, content endorsement – in newspapers, on social media – depends on “our own belief frameworks”. How could it depend on “web of beliefs” (Quine and Ullian 1970/2021) if what matters is attentiveness, inquisitiveness, discernment in appreciating evidence (etc.), most of which seem non-propositional agential properties? The point, we submit, is that belief frameworks shouldn’t be construed as webs of propositional beliefs, but should rather also include how update is performed (including different doxastic biases, in confirmation, memory, and so on), how context is identified, how situationally flexible or variable our

¹⁰ Lackey seems entirely correct in distinguishing an epistemological concern for the accuracy and reliability of the evidence we have available for our views, on the one hand, from the metaevaluative perspective upon our standing as potential epistemic agents, on the other hand. It may well be that epistemology is properly oriented primarily towards what Lackey identifies as its aims. The concern with epistemic autonomy regards an appreciation of agential propensities, capabilities and acts in response to the evidence available. It is from this standpoint that we explore the contours of epistemic agency in what follows; the issue is not one in epistemology but in metaepistemology.

beliefs might be,¹¹ and how much of that flexibility to bring to bear upon new content we interact with. Not as itemized attitudes (isolated and propositional) but rather than as outlook, view, perspective anchored in one's evidence-reaping behavior and one's communicative behavior.

8. Jaster and Lanius on truthfulness

We aim to sketch ways in which truth and evidence-related epistemic values (realized at varying degrees in different contexts) might bear upon the autonomy with which we communicate and with which we hold our doxastic attitudes. It seems to us that Romy Jaster and David Lanius propose a notion of truthfulness that is apt for the task. They write:

“When things go well, a news report is true—both in its literal content (“what is said”) and in its communicative content (“what it pragmatically conveys”). In the case of fake news, the literal or communicative content is false. Thus, fake news lacks truth on one of the two dimensions. An example for fake news with false literal content is the Pizzagate story that circulated in Russian media outlets in 2016.”⁶ (Jaster and Lanius, p. 21)

¹¹ A more general term, e.g. “doxastic stances”, would miss the point that these non-propositional agential dispositions are notwithstanding geared toward forming truth-apt beliefs. Moreover, we take such frameworks to include what we accept in addition to what we believe, if one wishes to contrast them (Elgin 2004). We don't propose this interpretation of “belief frameworks” as exegetical to Lackey – rather, we think pluralism about such notions is welcome, and only aim to sketch a notion lax enough at to accommodate our everyday practices of digital communication.

Leaving this particular example behind,¹² what Jaster and Lanius suggest with respect to a news report seems to be not just that there is a presumption to be had a news report is true, but also that a distinct presumption that releasing the news report was done in a truthful manner, i.e. with the communicative intention of spreading the truth. By our lights, key to Jaster and Lanius's view is the general distinct between truth and truthfulness. One might be truthful and yet speak falsely. One might also be mendacious in asserting¹³ what is nonetheless true.

The question is how to best understand truthfulness in this context; at least two options are open. On one construal, truthfulness presupposes intent. What one says is truthful to the extent that, in saying it, one intends to speak the truth. On this view, intending to speak the truth is a distinct mental state, with directly specifiable content. A different construal is also possible. Think of the honest, who speak the truth out of habit. They need not necessarily harbor an explicit intention to speak the truth as a discernible mental state, transparent to

¹² One red herring: the Pizzagate story might not be the best example; it is a blatant conspiracy theory that doesn't rely on a shred of evidence. In contrast, other conspiracy theories might invoke evidence in their favor, evidence which they skew and reinterpret to suit purposes alien to what the evidence would command if read head on.

¹³ We also obviate the important semantic question of whether what is said in context and what is pragmatically conveyed by a communicative act are always or at least sometimes distinct. The metaphysics of propositions and how they differ both from the literal content of the words used in an asserting that proposition and from the inferred pragmatic entailments, implicatures, new nuances or suggestions is a delicate matter best set to one side here. The metaphysics of thought and the metaepistemology of doxastic autonomy differ.

themselves in reflection. Nor need their truthfulness be transparent to their interlocutors. Being truthful out of habit is a disposition we attribute to the communicative agent regardless of the itemized intentions they may manifest as they speak.

Construals differ. Despite malevolent intentions, one might notwithstanding speak truly – out of habit. Contrarywise, an absent habit to speak the truth could be compensated for, and indeed gradually build on the basis of, specific communicative intentions to speak the truth. (We bracket the thorny case of inveterate truth-tellers who have no intention to speak the truth, nor is that a live issue for them, ever.) And ways of construing truthfulness aren't limited to assertion. One might truthfully ask a question, issue a command, make a promise or even being ironical, sarcastic or prone to *double-entendres*, regardless of whether these specific accuracy, felicity or correctness conditions for these respective speech acts are met or not.

Speaking the truth (now or habitually), reaping the right evidence and appreciating it appropriately are not just matters of more or less – they can be speaker-oriented or listener-oriented. The communication environment (social media, a newspaper, the public square, science journalism, etc) could be such as to promote these values or not. Careful speakers and listeners may be such as to promote these values or not. And all sorts of trade-offs are possible: careful speakers and careless interlocutors in an epistemically poor environment (virologists during pandemics); careless speakers in an epistemically noisy environment where truth and falsehoods mix (politicians on the campaign trail); proficient listeners in a carefully curated environment (a hybrid lecture filled with attentive

students who care about the subject), and so on. In what follows, we focus on a speaker-oriented view, but it bears mention this is only an option, and we take much of what we say below to apply, *mutatis mutandis*, to interlocutors and environments/media as well.

Let's connect truthfulness with Lackey's view (as interpreted above) the digital communicative acts presuppose doxasticevidence-curatorial abilities. If Jaster and Lanius are right, and if the foregoing does not misinterpret them, one natural suggestion is that a communicative act is properly performed, in a digital environment or otherwise, if agents performing that act are truthful, and doxasticone's abilities to appreciate evidence for and against the views that one either puts forward or criticizes are causal prerequisites for one's truthfulness as a *bona fide* communicative agent.

Taking stock, what we seem to be faced with is – *an array of different agential abilities: to intend to speak the truth in the here and now; to habitually speak the truth; to intend to let oneself be guided by evidence, to seek evidence so that it may become available, to appraise the evidence gathered for its significance in the context of the content that is to be communicated, to represent oneself as an agent that is sensitive to evidence, and to let that self-representation guide both one's evidence seeking activity, one's evaluative appraisal of the evidence gathered, and one's ensuing communicative acts.*

In what follows, we explore the line of thought that there need be no specific intrinsic content-based unity to these agential abilities. Rather, the next best that might bring together these otherwise disparate abilities is what might, in Lackey's vein, be called one's "belief framework"— or alternatively, if we follow virtue theorists, be deemed a core component of one's

intellectual character. (With the running *caveat* that this, too, is best construed not as narrowly epistemic but rather as broadly doxastic.)

This view of doxastic autonomy, as it begins to be fleshed out relative to communicative acts in digital environments, seems to fit well with Habermas' (1984) construal of the norms governing the rational appraisal of communicative acts. One key difference being that where Habermas takes regulation by these norms as constitutive of what it is for such acts to be deemed as communicative, my suggestion is that habitual conformity to the norms of truth, truthfulness, and sensitivity to evidence are often a part of one's approach to communication (in the ideal case), and often enough best construed as manifesting one's intellectual character-based autonomy (in the ideal case, again). And everyday users of social media and other digital communication technologies may depart from such ideals as much as feasible, turning virtue into neutral abilities, features or sliding into limitations and vices. As briefly sketched here, this is all *natural*. We make no claim that communicative doxastic autonomy is how it should be, nor that it is ideally rational, nor that communicative acts constituted by such performances are ideal, in any sense other than that of being presumed reasonable relative to the contingencies of the agents and circumstances that causally brought them about.

9. Hieronymi on responsibility for attitudes, Baehr on reliable epistemic character

At this point it is useful, we believe, to contrast the foregoing with two other views put forward and recent epistemology, Pamela Hieronymi and Jason Baehr. For Hieronymi:

“One’s commitment-constituted attitudes, then, embody one’s take on things, on whether some object is true, good, dangerous, offensive, worthwhile, rightly one’s own, to be brought about, etc. ...if being responsible is a matter of being both open to such assessments and (depending on the outcome of that assessment) the appropriate target of certain reactions, then the class of things for which one is most fundamentally responsible - the class of things that can constitute the object of that assessment and reaction - are also necessarily non-voluntary.” (Hieronymi, p. 371)

Hieronymi draws the contours of a general outlook with respect to our attitudes. Commitment, here, showing the kind of dispositional agential involvement that may speak not just to the attitude harbored by the agent but also to how that agent self-represents or self-constitutes as the holder of that attitude. This fits hand in glove with a neutral, general-purpose, characterization of “belief frameworks” as needed as a backdrop for how the agent appraises evidence in Lackey’s view.

Recall the newspaper. Why would one read the newspaper twice or thrice? Whatever it is, it must have something to do with the agent in question rather than with this specific copy of the newspaper now read. What, in an agent, might be responsive here? Perhaps attentiveness or conscientiousness? On the contrary, we contend agential abilities *need* not be specified further, beyond the sheer presumption their involvement is reasonable for an agent in those specific circumstances. This reasonability warrants us in imputing some *modicum* of commitment that the agent has (or should have) in reading the paper with some understanding of the news it delivers.

A narrower appraisal of what doxastic agency does can be gleaned from Jason Baehr's take on reliable epistemic character:

“According to virtue reliabilists, a personal quality is an intellectual virtue only if it plays a critical or salient role in getting a person to the truth, only if it best explains why a person reaches the truth... I argue that character virtues sometimes satisfy virtue reliabilists' formal conditions for an intellectual virtue. This point, together with the fact that virtue reliabilists generally view knowledge as (roughly) true belief arising from an exercise of intellectual virtue, reveals that intellectual character virtues are indeed relevant to virtue reliabilist accounts of knowledge.” (Baehr, p. 198)

For Baehr, the issue is finding common ground between responsibilist and reliabilist virtue epistemologies and, respectively, where such common ground cannot be achieved, what benefits each of the two families of views might have over the other, as well as whether a more encompassing and richer view of epistemic character should be in the offing. It is this latter goal that we are primarily concerned with here.

The target of evaluation, for Baehr, is not epistemically optimal or ideal behavior, which made best be appraised in terms of intellectual virtues and cognitive skills, but rather the kind of doxastically imperfect behavior that makes most sense and becomes reasonable only in the light of specified non-ideal circumstances, such as participation in an echo chamber or being exposed to fake news or unwittingly being part of a bubble that bounces conspiracy theories.

To this end, a broader understanding of one's doxastic autonomy is needed than what can be characterized

purely in terms of intellectual character. It is for this reason that we suggested, in the previous section, the constellation of cognitive abilities of appraising evidence and using it in informing one's autonomy in the delivery of communicative acts. That set of cognitive abilities might sometimes amount to skills, expertise, intellectual virtue or virtues, but need not always do so.

Attentiveness might be virtuous. But at what point does the newspaper reader stop being inattentive and start paying attention in a purely virtuous rather than only non-vicious way? Again, conscientiousness might be virtuous. But at what point does the newspaper reader that reads it over and over again start to manifest intellectual virtue rather than being subject to sunk costs in an epistemically sterile project? Perhaps mixing both, consider the ability to look for evidence for views one puts forward. If this is unaccompanied by one's ability to suitably recognize the evidence relevant to one's views, insisting upon looking for it when one is unable to find and recognize it seems less virtuous than vicious. Appreciating the exercise of a cognitive ability or disposition that an epistemic and pragmatic agent manifests as virtuous, rather than vicious or perhaps value-neutral, seems to be an entirely context-sensitive affair.

If what we said so far is on the right track, this picture seems to emerge: at the expense of a narrow focus on intellectual character, especially when it comes to non-ideal behavior, we would do best to think of a broader "belief framework". that includes abilities, habits, dispositions, bents and knacks. Such repeatable occurrences might underlie what Hieronymi calls commitment- constituted attitudes, with the rider that

what matters for us is the *natural* manifestation of one's doxastic autonomy (or the habits themselves rather than which rational norms they evince conformity to).

10. Alston on epistemic desiderata

The picture that just emerged meshes well with William Alston's project of replacing wholesale talk of "justification" with a more context sensitive approach to which aims of cognition, doxastic dimensions or epistemic *desiderata* might be at play in either supporting or undermining the views one intimates *via* communication. Alston writes:

"No one disputes that it is a good thing for the attainment of the aims of cognition that one have sufficient evidence for beliefs, that one conduct one's doxastic activities in such a way as not to violate intellectual obligations, and so on. This point will be the keystone of the new approach to the epistemology of belief developed and defended in this book. My proposal will be that we give up the attempt to determine just what is required for a belief to be "justified", indeed that we give up the supposition that there is any such property of beliefs that we need to get at the nature and conditions of, and instead focus on a variety of epistemic desiderata for beliefs, features of belief that are desirable from the epistemic point of view" (Alston 2005, p. 19)

Let's drive the point home. On top of all the abilities, virtues, skills, or sheer wittiness to exploit the circumstances garnered in communication, consider all these other possible epistemic *desiderata*:

- Having reasons in favor of a belief
- Varieties of epistemic statuses – online included
- Having reasons to spread a belief

- The belief in question favors epistemic flourishing (character, etc.)
- The belief in question furthers compliance to epistemic norms
- Reflective endorsement
- The belief in question is backed by evidence (*realiter*)
- The belief in question is backed by evidence (relevant by the agent's lights)
- The belief in question closely inferentially connects to other beliefs had by the Agent
- The belief in question closely inferentially connects to the agent's intentions, plans, practical deliberations
- The belief in question prompts the agent to attend more closely, thereby encouraging conscientiousness (personal-level) and salience of what is relevant (sub-personal level)

We hope you will agree, and do so as a matter of obviousness, that there are many more dimensions along which one's doxastic autonomy in holding, maintaining or altering one's views in communication might be assessed.

Conclusion

Our starting point has been Slobodan Divjak's exploration of the risk of spiraling into varied forms the digital totalitarianism. We followed Divjak in his emphasis on information control and on the opportunities and risks to alter or control information and communication that are afforded by progress and digital technologies underwriting social media. Naturally, this raises the question of autonomy for agents that aim to

know, believe or speak truly and truthfully. How far indeed communication is from such an ideal is transparent in the context of fake news and the spread of conspiracy theories.

Precisely in order to do justice to the variety of non-ideal doxastic practices and to the in- between of having full autonomy or no autonomy at all – we suggested it is best to move away from a metaepistemology focused on evidence appreciation alone, as well as from a metaepistemology focused on intellectual virtues or intellectual vices alone. Indeed, such metaepistemological *monisms* seem not always to capture the varied ways in which agents who might otherwise be considered reasonable (if all the information were in) would understandably be regarded as manifesting epistemically pernicious behavior in echo chambers across social media. *A multidimensional, pluralistic, context-sensitive, open-ended list of partially specified cognitive dimensions* along which to appraise the extent of one's doxastic autonomy in communication seems better advised.

Acknowledgments

In its current form, this text was heavily influenced by “The effects of LLM interaction in digital and virtual environments on TOM” research project, funded by the University of Bucharest as a grant for enhancing institutional performance, contract no. 10504/25.10.2024.

Many thanks to Mircea Dumitru and Alexandru Dragomir at the University of Bucharest for a 2022 project draft that anticipated this talk, both with important comments and suggestions. Other participants in that project proposal include Dana Deac, Constantin

Vică and Sorin Costreie. We also found a graduate conference from students in the “Mind the Brain” programme (UB Philosophy, especially Mihai Petrișor) extremely fruitful: “From echo chambers to context collapse” (June 2024, <https://philevents.org/event/show/123010>), where Sandra Brânzaru gave a talk titled “Social media and context (or narrative)-collapse” and first discussed many of the ideas presented here.

Thanks are also due to the organizers and audience at the “Capacities and the Good” workshop at Freie Universität Berlin (September 2022), where Andrei Mărășoiu presented “Vicious coordination, epistemic bubbles and the slide into conspiracy theories”. More info about that project here: <https://www.capacities-and-the-good.de/> As well as to the audience of the Interdisciplinary conference devoted to research results at the University of Bucharest (November 2024).

Finally, heartfelt thanks to Milenko Bodin and to the organizing team at the Appleid Philosophy and Social Research Center in Belgrade for a wonderful welcome and seamless conference in the “Felix Romuliana” 24 th International Summer School of Philosophy organized on 6-8 September 2024 in Zaječar, Serbia, where we presented this contribution. The conference topic, “Philosophy and Intelligence – Risks of Digital Totalitarianism”, is of notable significance today and deserves close scrutiny.

References

1. Alcoff, Linda. 2007. "Epistemologies of Ignorance: Three types". In: Sharon Sullivan & Nancy Tuana (eds.), *Race and Epistemologies of Ignorance* (pp. 39-58). Albany, NY: State University of New York Press.
2. Alston, William. 2005. *Beyond "Justification": Dimensions of Epistemic Evaluation*. Ithaca: Cornell University Press.
3. Baehr, Jason. 2006. "Character, Reliability and Virtue Epistemology". *Philosophical Quarterly* 56 (223): 193-212.
4. Brânzaru, Sandra Cătălina. 2024. "Posthumaning Creatively". *Revista de Filosofie Aplicată* 7, (12-13): 27-45.
<https://www.filosofieaplicata.ro/index.php/filap/article/download/142/102>
5. Buss, Sarah & Andrea Westlund. 2018. "Personal Autonomy". *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2018/entries/personal-autonomy/>
6. Chalmers, David. 2019. "The Virtual as the Digital". *Disputatio* 11 (55): 453-486.
7. Davis, Jenny L., and Nathan Jurgenson. 2014. "Context collapse: Theorizing context collusions and collisions." *Information, communication & society* 17, no. 4: 476-485.
8. Divjak, Slobodan. "Digital totalitarianism? Sophisticated technologies and the possibility of creation". *Keynote address at the "Felix Romuliana" summer school*, Zaječar, 2024.
9. Douglas, Karen M., Robbie M. Sutton, and Aleksandra Cichocka. 2017. "The psychology of conspiracy

- theories”. *Current directions in psychological science* 26 (6): 538-542.
10. Elgin, Catherine. 2004. “True enough”. *Philosophical Issues* 14 (1): 113-131.
 11. Frankfurt, Harry. 1971. “Freedom of the will and the concept of a person”. *Journal of Philosophy* 68 (1): 5-20.
 12. Fumerton, Richard. 1995. *Metaepistemology and Skepticism*, Lanham, MD: Rowman & Littlefield Publishers.
 13. Frost-Arnold, Karen. "The epistemic dangers of context collapse online." *Applied epistemology* (2021): 437-456.
 14. Goldberg, Sandy. 2011. “The Division of Epistemic Labor”. *Episteme* 8 (1) :112-125.
 15. Greco, John. 2014. “*Episteme*: Knowledge and Understanding”. In: K. Timpe & C.A. Boyd (eds.), *Virtues and Their Vices* (pp. 285-301). Oxford: Oxford University Press.
 16. Grice, H. Paul. 1957. “Meaning”. *Philosophical Review* 66 (3): 377-388.
 17. Habermas, Jürgen. 1984. *The theory of communicative action*. Boston: Beacon Press.
 18. Hieronymi, Pamela. 2008. “Responsibility for believing”. *Synthese* 161 (3): 357-373.
 19. Iftode Cristian, Alexandra Zorilă, Constantin Vică & Muriel Leuenberger. Forthcoming. “‘A Life of Our Own’: Why Authenticity is More Than a Condition for Autonomy”. *Journal of Value Inquiry* Online First 1-26.
 20. Jaster, Romy & Lanius, David. 2021. “Speaking of fake news”. In: *The epistemology of fake news*. Eds.

- Sven Bernecker, Amy Flowerree and Thomas Grundmann. Oxford University Press.
21. Korsgaard, Christine. 1996. "Reflective endorsement". In: *The sources of normativity*. Harvard University Press.
 22. Lackey, Jennifer. 2021. "Echo chambers, fake news, and social epistemology". In: *The epistemology of fake news*. Eds. Sven Bernecker, Amy Flowerree and Thomas Grundmann. Oxford University Press.
 23. Marwick, Alice E., & Boyd, Danah 2011. "I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience". *New media & society* 13(1): 114-133.
 24. Moran, Richard. 2012. *Authority and self-estrangement: an essay on self-knowledge*. Princeton: Princeton University Press.
 25. Quine, Willard van Orman & Joseph S. Ullian. 1970/2021. *The Web of Belief*. New York: Random House. Translated in Romanian as *Țesătura opiniilor*, transl. Mircea Dumitru. Iași: Polirom.
 26. Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. Transl. G.E.M. Anscombe. Oxford: Blackwell.