Sock Puppetry in Online Communication

GRACE PATERSON

Abstract

This paper concerns sock puppetry, a practice which involves an individual communicating under multiple pseudonymous identities in a manner that makes it seem as though these personas are distinct conversational participants. I provide a definition of sock puppetry that is more narrow than other definitions, allowing it to be distinguished from related phenomena. I then analyse some of the ways in which sock puppetry can interfere with social tools we use for establishing trust within an online community, evaluating a speaker's credibility, and generally deciding who and what to believe.

1. Introduction

In 2020, amidst the upheaval of COVID-19, the death of a prominent Native American and LGBTQ+ anthropologist was announced over Twitter[1] by her close friend BethAnn McLaughlin. At first, there was an outpouring of grief and support. The anthropologist, who went by the Twitter handle 'Sciencing_Bi', was an active and well-respected community member, participating in many online initiatives aimed at combating injustices within the academy. However, after a number of strange events, including an awkward Zoom memorial, it became apparent that Sciencing_Bi had never actually existed. She was, as it turned out, a particularly elaborate alter ego for McLaughlin herself.[2]

The juggling of accounts involved in McLaughlin's deceit marks it as a particularly striking instance of sock puppetry. *Sock puppetry*, as I shall define the term, is a practice where an individual presents

[1] The social media platform Twitter was acquired by Elon Musk in 2022, who changed its name to X. Since then, it has undergone, and continues to undergo, a series of significant and rapid changes in its functionality, policies, norms, userbase, and culture. In light of this instability, I have limited my discussion in this paper to the platform as it was prior to the Musk acquisition.

[2] https://www.nytimes.com/2020/08/04/style/college-coronavirus-hoax.html.

doi:10.1017/S0031819124000123

themselves under multiple guises which behave as if they were distinct people having an organic conversation. McLaughlin engaged in sock puppetry when she operated her fake account alongside an account in her own name and staged various online dialogues between her two personas, even going so far as to reference offline activities she and her alter ego were purported to have enjoyed together. It would not have been impossible for McLaughlin to create her false persona without engaging in sock puppetry, but it would have been more difficult. By performing interactions between herself and alter ego, she imbued the persona with a veneer of credibility and generally smoothed its entry into a community of peers.

This paper will sketch an account of sock puppetry and consider why it is useful to bad actors such as McLaughlin. More specifically, my aim is threefold: (1) to clarify what sock puppetry is and what makes it different from other kinds of related phenomena, (2) to identify how sock puppetry can have distorting effects on conversations, and (3) to explain some of the specific ways in which such distortions may be exploited by the puppeteer.

I will focus here on sock puppetry in online communication, although it is worth noting that it has some precedent in older mediums as well. In principle, sock puppets can be crafted anywhere speakers are able to present themselves under multiple different pseudonyms that are not obviously traceable to a single source. The American founding fathers, for instance, were famous for their use of pseudonyms in pamphlets and articles. Some of Alexander Hamilton's antics seem not so different from how sock puppetry is sometimes practiced today:

> Hamilton was not content to write as Camillus alone. Two days after his second essay appeared, he began to publish, in the same paper, a parallel series as "Philo Camillus." For several weeks, Philo Camillus indulged in extravagant praise of Camillus and kept up a running attack on their Republican adversaries. The prolific Hamilton was now writing pseudonymous commentaries on his own pseudonymous essays. (Chernow, 2016, p. 494)

While sock puppetry is not, therefore, exclusive to online communication, online forums have greatly reduced the barrier of entry for aspiring puppeteers. These days, you do not need to be publishing in magazines or pamphlets, as Hamilton and his peers were doing; you simply need to be able to make for yourself more than one account in the same space. As such, sock puppetry has emerged as a phenomenon with which individuals, online communities, and social media platforms must all contend.

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

The remainder of the paper will proceed as follows: in §2 I will discuss how individuals may use a number of different kinds of pseudonymity in online spaces. Among these, I will isolate sock puppetry as being of particular interest and provide a definition that makes clear how it differs from these other phenomena. In §3 I will describe some important ways in which online communication differs from face-to-face communication. I will argue that these differences result in important changes to how we come to trust one another, and in particular that there is a heavier reliance online on specific social indicators of trustworthiness. This creates a fertile environment for sock puppetry. In §4 I will describe specific exploits that rely on the use of sock puppets to target the vulnerabilities discussed in §3. In particular, I will discuss a form of credibility bootstrapping where an individual can gain the trust of members of the community by having their own alter egos perform the trust they seek from others.

2. Dramatis Personae

The term 'sock puppet' in reference to personas (rather than literal puppets made of socks) is a relatively old piece of internet vernacular, traceable to the early 90s USENET (Levine, 2014, p. 854). Today, it is common for online communication spaces to have some kind of explicit policy aimed at mitigating problems with sock puppets, although the strategies employed vary widely. Some spaces have fairly strict (though not exceptionless) 'one user one account' policies (Facebook, Wikipedia), while others opt to ban specific ways of using sock puppets such as to spread spam or inflate engagement metrics (Twitter, Reddit, TikTok).

That said, there is still a fair bit of variability in how precisely the term is used.[3] Some researchers identify sock puppets with misrepresentations of the puppeteer, defining them as false identities (see e.g., Levine, 2014; Greyson and Costello, 2021). Others emphasise the multiplicity of personas involved, defining a sock puppet as 'a user account controlled by an individual who has at least one other account' (Kumar *et al.*, 2017, p. 858). These definitions have the virtue of simplicity and, at least in the latter case, ease of operationalising for purposes such as the algorithmic detection of suspected sock puppet accounts in specific online spaces such as Wikipedia.

[3] Not to mention variability in how it is spelled, with 'sockpuppet', 'sock puppet', and 'sock-puppet' all in circulation.

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

However, in collapsing sock puppetry with other forms of online deception, such definitions admit what I would consider false positives (for instance, accounts purporting to be written by cats and dogs) while also obscuring those features that are unique to sock puppetry.

Since my aim is to examine some of what is special, and potentially problematic, about sock puppets qua sock puppets, it will be helpful to define a sock puppet in a more narrow way:

> *Sock puppet (sock)*
> One of multiple pseudonymous identities employed by an individual in a particular conversational or social context with the expectation that others will believe (falsely) that these represent distinct participants.

Sock puppets can be distinguished from the following related and frequently overlapping phenomena:

> *Alternative accounts (alts)*
> Pseudonymous identities used by the same individual either openly (with their connections obvious) or in different contexts. Alts are used for a range of purposes including humour, self-expression, identity exploration, protecting one's privacy, and keeping one's social roles separate (for instance, maintaining both a work account and a personal account on social media).

> *Fake identities (fakes)*
> Pseudonymous identities which actively misrepresent socially important facts about the individual behind them. Kinds of fakes include impersonations of specific individuals (e.g., public figures or celebrities), as well as cases of 'digital blackface' (when a non-black person assumes a fake black identity). When these identities are used to cultivate close relationships with others it is sometimes referred to as 'catfishing'.

> *Bots*
> These are automated accounts, often made and controlled *en masse*.

Sock puppets, alts, and fakes, and (at least some) bots are all forms of *pseudonymous speech*: a kind of *partial anonymity* where the audience is in the position to recognise that the same individual is behind everything said using the pseudonym, but is preventing from recognising that same individual in other contexts, for instance on the street or when they communicate under a different pseudonym (Paterson, 2020). Going forward, let us call the pseudonymous identities created through such speech personas. Consistent use of a

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

pseudonym over a long period of time can allow a speaker to construct a remarkably thick persona – one which has a complex history, reputation, and set of social bonds within a particular community. However, this is not essential; some personas are introduced only to make brief conversational interventions and, as such, remain quite thin.

Notice that the above forms of pseudonymous speech are tightly related and often co-occur. To give an example, when in 2020 Dean Browning (a white American congressman) Tweeted 'I'm a black gay guy and I can personally say that Obama did nothing for me, …' from his official government account in reply to another user disputing a point he had made, he was clumsily using a *fake* persona as a *sock puppet*. The persona was *fake* because he was not, in fact, a 'black gay guy', and it was a *sock puppet* because he was using it to reply to one of his other accounts. If he had kept his alter ego properly distinct from his main account, by not interacting with or commenting on himself, it still would have been fake, but an alt rather than a sock.[4]

Of course, a persona can be a fake without being either an alt or a sock-puppet if it is the only persona actively in use by the individual. Multiple personas can likewise be alts without being sock puppets if the user employing them makes explicit that they are controlled by the same individual or, again, simply avoids having them interact with one another. Finally, the same points apply to bots, which should be considered a form of sock puppet when employed to, for instance, inflate engagement numbers or flood reviews for some product,[5] but not when they simply do such things as generate an amusing social media post once an hour or periodically remind participants in a chat room about the moderation rules currently in effect.

The essential characteristic of sock puppetry as opposed to having alts is in a pattern of usage which allows for personas to interact with one another in ways that obscure their relationship to one another. Because of this, sock puppetry introduces potential misapprehensions into the background of a conversation, in particular related to how many people are involved in a particular exchange. This kind

[4] The full story gets more convoluted: https://www.nytimes.com/2020/11/10/style/dean-browning-patti-labelle.html. Browning's apparent mistake exemplifies one of the ways sock puppets are sometimes uncovered: the puppeteer gets their various identities confused.

[5] It is worth flagging that the ongoing development of conversationally sophisticated AI creates the potential for the use of bots in much more interactive forms of sock puppetry than the kinds I have mentioned here.

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

of misinformation sounds benign but can, as we shall see, have significant consequences for how other conversational participants interpret what has been said.

Since the puppeteer is aware that they may be creating false beliefs in others, sock puppetry is, to at least some degree, inherently deceptive. That said, it needn't always be malicious or harmful. Consider a case where an individual has one persona that they use for casual conversations and a different one that is used exclusively to act as a moderator. These personas could occasionally appear in the same conversations in a way that they realise could make it unclear to at least some participants that it is the same person behind both. The puppeteer may not, in other words, intend to create such a false belief, but nonetheless recognise the potential for it to arise as a side-effect of how they are juggling personas. In some contexts, where the stakes are low or the interaction brief, it may simply not make sense to take the time to explicitly flag the accounts as sock puppets. In other contexts, however, taking preventative measures against this kind of unintended sock puppetry might be warranted. Thus, under current Wikipedia policy, certain kinds of alternative accounts are allowed with the caveat that '[u]nless when doing so would defeat the purpose of having a legitimate alternative account, editors using alternative accounts should provide links between the accounts.'[6]

It is possible (although perhaps not common) for a person to unknowingly employ multiple personas in a way that creates the impression that they are distinct individuals. Imagine, for instance, responding to a decades-old forum post that, as it turns out, your younger self wrote using an account you have forgotten about. On the definition here, this is not sock puppetry because you do not anticipate creating a false belief in others (or yourself, as it so happens). At the same time, an individual who genuinely intends for such misunderstandings to arise is engaging in sock puppetry even if nobody is actually fooled. Incompetent sock puppetry is still sock puppetry.

Before moving on, there is one more phenomenon worth flagging because of its interesting parallels and connections to sock puppetry. This is the use of *collective pseudonyms*. Where sock puppetry involves one person constructing many personas, collective pseudonymity involves many people constructing and presenting themselves together under the guise of a group persona. The mathematical collective Bourbaki is a famous example, as is the online hacktivist group that

[6] https://en.wikipedia.org/wiki/Wikipedia:Sockpuppetry#Alternative_account_notification.

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

goes by Anonymous. Mixed cases are possible here too. For instance, in 2015, 381 Wikipedia accounts were banned for being part of a for-profit editing ring. Here a group of people coordinated around puppeteering hundreds of accounts that both wrote and gave apparent legitimacy to various articles. These individuals all made use of multiple sock puppet accounts; and most of those accounts were also collectively manipulated by multiple members of the group.[7] As such, the accounts were both group personas and sock puppets.

3. Understanding the Puppet Theatre

Sock puppets, by their nature, obfuscate information about the discursive context in which they operate – specifically, information about who is present, and how these participants are related to one another. This information is not incidental or idle; it is essential to our ability to competently navigate conversations and relationships with one another. In this section, we will discuss the role this background information plays in online communication and why, therefore, sock puppetry can be more of a hazard in these contexts than in some others.

Of particular importance here is the manner in which sock puppetry affects *trust*, understood here as a three-place relation in which *one person* trusts *another* within a particular *domain* (Jones, 1996). So, for instance, I might trust one person in my network to supply reliable information about the war in Ukraine and another to act as a fair and effective moderator, but not vice versa. When deciding whether to trust someone with something, we must judge both whether they are competent with regard to the domain, and whether they are sufficiently well meaning. One individual may be competent but malicious, and another benevolent but incompetent. Thus, a self-interested doctor may know very well that they are recommending a treatment for COVID-19 that does not work, but choose to do so anyway because it makes them money. One of their followers might recommend the same treatment because they genuinely believe that it works and lack the knowledge required to properly fact-check the

[7] See Wikimedia's statement from the time at: https://diff.wikimedia.org/2015/08/31/wikipedia-accounts-blocked-paid-advocacy/. The case was widely reported upon and is also extensively documented on Wikipedia itself (as, for that matter, are all sock puppetry investigations on the platform).

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

rumours. Both these individuals are untrustworthy, albeit in different ways.

The problem of evaluating competing claims about matters of fact is obviously not unique to those claims made online. Nonetheless, the many ways which the internet has expanded information access and reach have introduced new complications to the mix. Alfano and Sullivan (2021) give a concise explanation of this problem:

> The internet has made available an unprecedented number of accurate sources. However, they must be sifted from the spammers, trolls, sealions, practical jokers, conspiracy theorists, counterintelligence sock-puppets, liars, and ordinary uninformed and misinformed citizens who also proliferate online. (Alfano and Sullivan, 2021, p. 482)

Online we have access to abundance of good information and expertise, but it has been mixed in with a seemingly equal abundance of misleading or false information. We are left to 'sift' through these masses of information, identifying the good amongst the bad. A reasonable strategy is to focus on the sources of information, placing our epistemic trust in those whom we deem trustworthy informants. Thus well-placed trust plays a particularly important role in our online epistemic lives.

Note, however, that the challenge here is not purely epistemic. It pertains to interpersonal trust more generally. Competently and responsibly navigating the online social landscape demands that individuals do more than simply sort good information (and informants) from bad, because malicious actors do more than simply spread false information. Online spaces have become launching points for, and central sites of, ongoing interpersonal relationships of all kinds. On a one-to-one basis, individuals develop online friendships, romances, and professional collaborations with one another. Together, they form, join, and expand communities ranging from fandoms and hobby groups, to support groups for chronic health conditions, to groups devoted to activism and political organisation.

In other words, people do not just *peruse information* online, they *meet other people* too. The ways in which individuals come to trust one another online are, as such, not limited to the epistemic trust that is required for believing one another's testimony. Those commiserating over shared struggles may, for instance, trust one another to maintain confidentiality while those asking for help may be trusting others to provide safe links to reputable resources. Whether it occurs online or off, individuals who trust in these ways

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

make themselves vulnerable to potentially serious harm in the event of a betrayal.

Determining whom to trust and what to believe is made more challenging by the fact that we do not have access to the same contextual information online as we do in offline face-to-face conversations. For instance, offline, we use non-linguistic cues such as eye contact and body language both in interpreting speakers' meaning and in making judgements of trustworthiness.[8] Online, we do not have access to these information sources and so cannot use them in making those judgments.[9]

We cannot, in other words, simply assume that individuals use the same communicative and interpretive strategies online as off. At the same time, the fact that there are *differences* between offline and online communicative contexts does not necessarily entail that online communicative are spaces less socially complex than offline spaces. This is not a simplistic story of one medium being informationally rich and another impoverished. As Scott (2022) observes, '[i]n face-to-face communication we can augment our messages with these extra-linguistic clues to our intended meaning. When we move online, other resources become available to us' (p. 15).

By making use of these alternative resources, individuals have found new ways of communicating nuance and meaning. Thus, while they may not be able to express themselves through body language, gestures, gaze, and intonation as they do in face-to-face encounters, they gain access to other expressive tools – for instance, novel lexical devices in the form of hashtags and emojis. New expressive resources naturally generate new interpretive tools. Judgements about character and motivation are therefore aided by looking at how speakers have employed these expressive tools.

Additionally, people have developed methods and tools specifically suited to the problem of evaluating credibility and trustworthiness in online spaces. Here two kinds of information have emerged as particularly useful: (i) a speaker's credentials, and (ii) the reactions of others to her and what she has said – in other words, what others in our network think of her. I will refer to strategies that make use of

[8] We should not, of course, assume that these heuristics are universally reliable. Indeed many of these sorts of cognitive short cuts are plausibly subject to pernicious biases related to aspects of the speaker's identity such as gender or race.

[9] Indeed, and perhaps counterintuitively, unavailability of many non-linguistic cues for the purposes of communication and interpretation appears to extend even to video-based communication (Scott, 2022, p. 15).

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

these pieces of information as credentialing and credibility crowd-sourcing respectively. Let us consider them more closely.

Credentialing involves checking to see if people are qualified to speak on a particular topic or perform a particular task – in other words, whether they satisfy the competence criterion for trustworthiness. Credentialing is especially important when the domain in question is one that involves specialised training, experience, or knowledge. Thus it most obviously comes up when we are evaluating claims made by those presenting themselves as experts or those speaking on technical topics. In these cases, we may check that the supposed expert has relevant degrees, titles, or certifications. Doing so involves us depending on an institution to supply a verification of the person's reliability with regard to the topic at hand. All things equal, it is better to trust someone with a medical degree from a reputable school to give medical advice than someone without one.

Perhaps less obviously, credentialing may be used in cases where the speaker is purporting to speak from a position of first personal experience such as may arise from membership in a particular community, or living in a certain country. Here, in the absence of institutional vetting, we may base our judgements on the person's available biographical information and the testimony of others who know them.

Credentialing of any type requires two main ingredients: access to sources of corroborating information pertaining to the speaker's credentials, and enough accurate identifying information about them to be able to actually locate and recognise the relevant corroborating information. In terms of access to potential corroborating sources, the internet has been a great boon. However, when it comes to access to the identifying information required to actually make use of these sources, we run into more difficulties. If, for instance, someone claims on an anonymous forum that they are a doctor, we have no way of verifying this. Thus the process of credentialing can be undercut by the anonymous and semi-anonymous character of many online social spaces. We are often, in these forums, in the position of having to evaluate claims made by individuals who are strangers or near strangers and are equipped with only the information that person has chosen to share. The upshot is that credentialing is a mixed bag online: we have more resources for researching an individual's background, but bad actors are also more able to manipulate their self-presentation in ways that confound this process.

Let us consider now the second heuristic mentioned above. Credibility crowd-sourcing involves us making use of social metrics to judge another's credibility and trustworthiness. For instance,

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

Boyd (2022) argues that one of the most salient markers of trustworthiness for scientific experts online is what he called genuine endorsement.[10] Specifically:

> The extent to which information presented by an expert is genuinely endorsed – i.e. endorsed by individuals because they think it is true – the more trustworthy the expert presenting it. (Boyd, 2022, p. 14)

That is, we consider the degree to which other people in our social circle believe and endorse claims made by experts we are less familiar with. More generally, we make use of signs and metrics that reflect how a particular speaker or speech act is received by the wider community, essentially checking to see to what extent the speaker is respected by others before deciding to trust them ourselves. Do their points get picked up and repeated by our peers? or are they scorned and mocked? Consider how vigorous nodding or sceptical scowling by friends around a dinner table might colour our reception of an unfamiliar speaker – that is a consequence of this form of crowd-sourcing.

Online, such social signifiers seem especially potent, such that 'people are inclined to believe information and sources if others do so also, without much scrutiny of the site content or source' (Metzger and Flanagin, 2013, p. 215) (see also Metzger *et al.*, 2010).

Moreover, measures of social reception have become a core part of the design of many forms of social media. Thus, online spaces are generally equipped with, or develop, new ways of tracking information that are intended to help users assess the speakers' reputation within the relevant community, as well as the overall reliability of whatever it is that they have said, often by using crowd-sourced metrics. The popular forum site Reddit, for instance, allows users to upvote and downvote posts based on their perceived quality, and this both affects the visibility of the post as well as granting the poster themselves more or less 'karma'. This points system is intended to track reputation so that 'your karma is a reflection of how much your contributions mean to the community'.[11] Karma is

[10] The other marker of trustworthiness Boyd considers is genuine cooperation and has to do with how a scientific expert presents information to their audience. The trustworthy expert 'takes into consideration the relevant beliefs and reasons likely possessed by an audience with the aim of having them believe truths' (Boyd, 2022, p. 14). In other words, they have both rhetorical skill and good intentions.

[11] https://support.reddithelp.com/hc/en-us/articles/204511829-What-is-karma.

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

shown in a user's public profile and so, in theory, should allow one to see at a glance if an individual has a history of contributing positively or negatively to the community.

The assumption behind the use of these tools is that what other members of the community think of a particular speaker or speech act can to some degree be gleaned by looking at forms of engagement that are explicitly aggregated, counted, and otherwise displayed. Interactions between users that take the form of likes, reposts, reaction emojis, votes, star ratings, and, of course, actual replies, all serve to show you what other people think of their target. So, for better or worse, a background assumption that we can trust those whom others already trust, and likewise that we should be suspicious of those that others already doubt, is hardwired into many of our online interactions.

Interestingly, users are adept at finding such social markers even when these are not explicitly designed for, and, moreover, at interpreting those that are part of the design in more sophisticated ways than their designers initially intended. For instance, on Twitter, the ratio between the number of interactions that tend to be supportive (likes, retweets) and the number of interactions that are often unsupportive or critical (quote tweets, comments) has become known as a way to quickly assess how the tweet was received by the community. A tweet with many comments and few likes is perceived as very unpopular and therefore may be viewed with suspicion or derision. This is an improvised metric devised by the community, rather than by the site administrators themselves, but it serves the same end of giving the audience a rough way to evaluate the credibility of claims even when the speaker is unfamiliar to them.

4. Puppet Shows

We have seen that establishing trust online brings with it special challenges, and that social indicators such as endorsement and agreement within a community are particularly important resources for evaluating the trustworthiness of others. When individuals lack either an in-person acquaintance with someone new or robust access to external credentials, they look to how other members of the community assess the speaker and her words.

One of the reasons sock puppetry is of interest is that it can directly interfere with such approaches. To see how this is, we will consider one particularly insidious kind of exploit that is sometimes used by malicious puppeteers. This is a kind of trust or credibility

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

bootstrapping and involves use of different personas to feign particular kinds of social uptake, thereby inflating (or deflating) the apparent credibility of whomever the puppeteer chooses (for instance, a preferred sock).

One form of this behaviour involves one persona explicitly endorsing or vouching for one another. This can enable the puppeteer to gradually infiltrate a community, accruing trust based on a foundation that is, at bottom, largely fictitious. Indeed, the long running ruse in our opening example involved several moves of this kind. With her primary account, McLaughlin was able to get people within her online community to trust her fictitious persona. Moreover, that the persona was an actual person and trustworthy, despite their being previously unknown to anyone else, was repeatedly attested to by reports of supposed offline interactions between the persona and McLaughlin herself. Once the persona was embedded within the community, McLaughlin was able to close the loop, exploiting the persona to present herself now as a trusted ally to Native Americans – despite being nothing of the kind.

McLaughlin is far from alone in practicing this form of deception. Indeed, a study of online discussions related to cases of Munchausen by Internet (MbI) – a condition in which an individual fakes being ill online in order to gain sympathy or attention – found at least some community level awareness of such exploits:

> [cues of trustworthiness] included meeting the person face-to-face or another participant vouching for their credibility. Both however were viewed as fallible, individuals have been known to act out MbI in real life and those vouching for the participant could be sock puppets. (Lawlor and Kirakowski, 2017, p. 108)

Recall that the practice of sock puppetry predates the internet – the internet simply introduced more and easier opportunities for its use. This point holds of these sorts of credibility bootstrapping as well. For instance, Barany (2020) documents how in the mid-twentieth century the mathematician Kosambi, having 'tried and failed to convince friends and colleagues at home and abroad that he had proven the Riemann Hypothesis' (Barany, 2020, p. 19) wielded what remained of his good standing within the mathematical community to get his proofs published under the guise of his persona Sven Ducray:

> Helped along by Kosambi's promotion, the papers received ordinary notice in both Mathematical Reviews and the Zentralblatt für Mathematik, with reviews signaling apparent

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

> errors. Representing himself as Ducray's mentor, Kosambi corresponded with at least one publisher on Ducray's behalf and traded on his own prestige to see Ducray's paper to print. (Barany, 2020, p. 19)

While Kosambi's own mathematical credibility was under strain, he was nonetheless still trusted (at least initially) with introducing new members to the mathematical community. As a result, he still had enough credibility to vouch for his fictitional 'mentee' as a serious new talent on the scene.

Let us unpack in more detail how and why such methods work. I take *vouching* to be a speech act that facilitates the spread of trust by harnessing existing trusting relationships to generate new ones Paterson (2022). Specifically, one person (here Kosambi), addressing someone who trusts them (colleagues, publishers), can vouch for a third party (Ducray) as trustworthy within a particular domain (mathematics). Doing this involves a certain taking of responsibility for what happens in virtue of the addressee trusting the vouchee. If the new person turns out to not, in fact, be trustworthy, whoever vouched for them can be called to account. The voucher thus makes herself vulnerable insofar as a betrayal of the addressee by the vouchee will be costly for herself as well.

This mechanism works because the person doing the vouching is able to act as a kind of insurance against betrayals by whomever they vouch for. The foundation for trust vouching offers therefore collapses if the person vouching is the same as the person being vouched for as happens when a puppeteer vouches for their own sock to a stranger. There are several reasons for this. Most obviously, there is no interpersonal trust relation between the voucher and vouchee, which is a precondition for vouching in the first place. But more importantly, if the sock betrays those who trust them, the puppeteer does not thereby incur any kind of cost. After all, they are the one doing the betraying. There is not a genuine sharing of risk between voucher and addressee, only the pretence of doing so. The puppeteers in these cases can induce trust in their victims by creating the appearance that a sock puppet has been properly vouched for. But the trust created has no grounds at all.

Now in the case of someone like Ducray the stakes are not especially high: the main problem is that the community has wasted their time revisiting mathematical 'proofs' that they have already concluded are invalid. But the harms are much greater when the act of vouching is used to impersonate a member of an identity group (as with McLaughlin), extract emotional labour from well-meaning strangers,

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

or just generally to embed oneself within a private community (see, for instance, cases documented in Greyson and Costello, 2021; Lawlor and Kirakowski, 2017).

An important note here is that the addressee does not need to trust the voucher *in the same way* or for the same thing as she does the vouchee. Seeing my anxiety about finding a new doctor, you might be able to vouch for your own GP, enabling me to make the leap in trusting them. But this does not require me to trust you to be my doctor. So if a puppeteer can convince community members to trust her judgement about other people sufficiently, she can create trust in one of her own puppets with respect to an entirely different domain. Thus McLaughlin harnessed trust in herself as a member of one community to vouch for her persona, who was (she claimed) a member of another community to which McLaughlin did not herself belong. And with the help of elaborate back and forth dialogue, she was equally able to boost trust in her 'real' persona as well.

5. Conclusion

We have seen that sock puppetry makes possible certain harmful exploits. Moreover, while successfully exposing bad actors can sometimes solve the immediate problem that they pose, it may come at the cost of longer-term degradation of trust within a community, as well as widespread changes in its membership and internal dynamics (Greyson and Costello, 2021, pp. 14–15). So the damage done by sock puppetry can outlast the actual presence of the puppets in the community.

In light of these points, it would be easy to conclude that we should take aggressive precautions against the possibility of sock puppetry in our communities. It is therefore worth bearing in mind that individuals having and maintaining multiple personas online (or off) is not in itself a problem and, indeed, can bring many benefits. There are numerous reasons why allowing the use of pseudonyms and alts on social media platforms remains in most cases harmless, and in many cases valuable and worthy of protection. We have all grown fairly adept at handling a rather high degree of anonymity and uncertainty in our online interactions.

But full identifiability is not required to prevent those exploits we considered here. This is because sock puppetry depends on the absence (or manipulation) of highly specific background information. That is, sock puppetry can arise wherever there is misinformation about what pseudonyms in use are actually co-referring, representing

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

therefore the same conversational participant under different guises. Exploits arise when such personas *interact with one another* in a manner that obfuscates the fact that they are being controlled by the same person.

A broader lesson we can take from what we have seen is that in social interactions we depend on having access to certain background facts about the structure of the conversation and social space, and some of these facts are very basic things indeed. Even in contexts with fairly high levels of anonymity, we rely on knowing, in some coarse-grained sense, who we are speaking to and what each distinct participant has said. Problems with sock puppetry are more acute online than off because in face-to-face conversations we can see the agent from whom particular speech acts are issuing. Absent skilled ventriloquists in the room, it is therefore easy to determine when two speech acts have the same source. Of course, there is still sometimes mischief that occurs between multiple face-to-face conversations, as when Lois Lane talks to Superman on Monday and Clark Kent on Tuesday. But we do not generally encounter these confusions within a single face-to-face conversation, whereas we very well may online.

Plausibly, then, a person's ability to successfully and, indeed, responsibly navigate online social spaces relies on them having a solid grasp of that space's structural features. For instance, speaking of social epistemic networks, Alfano (2021) suggests that 'there are is [sic] a virtue associated with monitoring and being disposed to rewire a network in which one occupies a receiver role, for instance by finding new sources or by cutting off sources one no longer trusts' (p. 8437). If that is correct, then practices such as sock puppetry represent a significant obstacle; for the effective exercise of such a virtue is directly undermined if the receiver is misled about what nodes are in the network itself. As such, virtuous individuals, so described, must also be attuned to the possibility that they may be under misapprehensions about their own community's network structure and membership. When monitoring their social network they should look for groups of nodes that present themselves as distinct but in fact are identical.

It is reasonable, I think, to expect that over time the ever-growing experience and sophistication of online communicators will help mitigate some of the harmful effects of sock puppetry. Some of our vulnerability to these behaviours reflect how, even after several decades, we are still in the process of developing norms and heuristics properly suited to the various online social settings in which we find ourselves. We should not abandon valuable aspects of online

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press

communication such as pseudonymity, nor should we give up on our time-tested methods of spreading trust through practices such as vouching. Instead we should work at mitigating pernicious forms of self-interaction.

Acknowledgements

Early versions of this paper were presented at the 2022 Social Ontology conference at the University of Vienna, and at a 2022 workshop on Online Speech at the University of London. My thanks to audiences at both events for their helpful feedback. I am also indebted to Anna Drożdżowicz, Andrew Tedder, and a very constructive reviewer for helpful comments. Finally, my thanks to the editors of this volume for their hard work and patience.

This project received funding from the Norges Forskningsråd (grant number 324393), and from the European Research Council (ERC) under the European Union Horizon 2020 research and innovation programme (grant agreement No. 740922).

References

Mark Alfano, 'Virtues for Agents in Directed Social Networks', *Synthese*, 199:3 (2021), 8423–42.

Mark Alfano and Emily Sullivan, 'Online Trust and Distrust', in Michael Hannon and Jeroen de Ridder (eds.), *The Routledge Handbook of Political Epistemology* (Abingdon: Routledge, 2021), 480–91.

Michael J. Barany, 'Impersonation and Personification in Mid-Twentieth Century Mathematics', *History of Science*, 58:4 (2020), 417–36.

Kenneth Boyd, 'Trusting Scientific Experts in an Online World', *Synthese*, 200:1 (2022), 1–21.

Ron Chernow, *Alexander Hamilto*n (London: Head of Zeus, 2016).

Devon Greyson and Kaitlin L. Costello: '"Emotional Strip-Mining": Sympathy Sockpuppets in Online Communities', *New Media & Society*, 25:12 (2021), 1–22.

Karen Jones, 'Trust as an Affective Attitude', *Ethics*, 107:1 (1996), 4–25.

Srijan Kumar, Justin Cheng, Jure Leskovec, and V.S. Subrahmanian, 'An Army of Me: Sockpuppets in Online Discussion Communities', in *Proceedings of the 26th International Conference on World Wide Web* (2017), 857–66.

Aideen Lawlor and Jurek Kirakowski, 'Claiming Someone Else's Pain: A Grounded Theory Analysis of Online Community Participants Experiences of Munchausen by Internet', *Computers in Human Behavior*, 74 (2017), 101–11.

Timothy R. Levine, *Encyclopaedia of Deception*, volume 2 (London: Sage Publications, 2014).
Miriam J. Metzger and Andrew J. Flanagin, 'Credibility and Trust of Information in Online Environments: The Use of Cognitive Heuristics', *Journal of Pragmatics*, 59 (2013), 210–20.
Miriam J. Metzger, Andrew J. Flanagin, and Ryan B Medders, 'Social and Heuristic Approaches to Credibility Evaluation Online', *Journal of Communication*, 60:3 (2010), 413–39.
Grace Paterson, 'Sincerely, Anonymous', *Thought: A Journal of Philosophy*, 9:3 (2020), 167–76.
Grace Paterson, 'Trusting on Another's Say-So', *Ergo*, 8:43 (2022), 520–37.
Kate Scott, *Pragmatics Online* (Abingdon: Routledge, 2022).

GRACE PATERSON (grace.paterson@inn.no) *is a philosopher of Language, Social Ontology and Action. She received her PhD from Stanford University in 2018 after which she was a postdoctoral researcher in Vienna, Austria for several years. She is currently a postdoctoral researcher at the Innland Norway University of Applied Sciences in Lillehammer, Norway.*

https://doi.org/10.1017/S0031819124000123 Published online by Cambridge University Press