Trustworthiness and Truth: The Epistemic Pitfalls of Internet Accountability

Karen Frost-Arnold[1]

frost-arnold@hws.edu

Abstract

Since anonymous agents can spread misinformation with impunity, many people advocate for greater accountability for internet speech. This paper provides a veritistic argument that accountability mechanisms can cause significant epistemic problems for internet encyclopedias and social media communities. I show that accountability mechanisms can undermine both the dissemination of true beliefs and the detection of error. Drawing on social psychology and behavioral economics, I suggest alternative mechanisms for increasing the trustworthiness of internet communication.

Introduction

Internet anonymity can encourage people to act in bad faith. For epistemologists, this raises concerns about the trustworthiness of internet communication. If anonymity makes it easy for agents to lie, makes it difficult for audiences to judge the competence of speakers, and prevents us from providing real-world punishments for those who undermine epistemic practices, then, so this concern goes, the internet provides a poor medium for the production and dissemination of knowledge. *Wikipedia* vandalism, Twitter pranks, and hoax blogs are now

[1] I thank P. D. Magnus, Ben Almassi, the works-in-progress group at Hobart & William Smith, and an anonymous reviewer for *Episteme* for helpful comments on earlier versions of this paper.

familiar features of our online lives. Recent cases of a corporation removing unflattering information from its *Wikipedia* pages, a Twitterer spreading the false rumor of flooding in the New York Stock Exchange during hurricane Sandy, and a fake blog on the Syrian uprising all heighten worries about the epistemic merit of internet communication (Borland 2007; Bell & Flock 2011; Kaczynski 2012).

A common response to these abuses maintains that the early ideals of a free and open internet were naïve—now that we see the tendency for abuse, mechanisms of accountability are necessary to ensure the epistemic value of internet communities. Some critics advocate for criminal and civil penalties for Twitterers who spread misinformation (Kaczynski 2012; Keen 2012). Others propose publicly naming and shaming *Wikipedia* vandals (Borland 2007). In addition, Alvin Goldman suggests that since many bloggers are not accountable to employers, blogging has less epistemic value than traditional journalism, in which journalists are motivated to keep their jobs (Goldman 2008: 121). These responses share the assumption that accountability is what is lacking, i.e., stronger systems of accountability would enhance the internet's epistemic status. This paper challenges this assumption.

While the benefits of accountability mechanisms are often discussed, their problems are largely ignored in social epistemology. I argue that there are several veritistic problems with internet accountability mechanisms. But my claim is not that accountability has no epistemic value. Much depends on the particularities of online communities. Further empirical research is needed on the effects of various accountability mechanisms. This paper lays out veritistic considerations that should be investigated before an accountability approach is advocated. I show that accountability mechanisms can both damage epistemic communities' ability to disseminate true beliefs and undermine their ability to weed out error.

Alternatives to accountability mechanisms are also neglected in social epistemology. Recent studies in social psychology and behavioral economics not only challenge the basic assumptions of the *Homo economicus* model behind accountability approaches to trustworthiness, but these experimental results also suggest alternative methods for promoting trustworthy behavior. Drawing on these methods for engaging users' pro-social and ethical motivations, I propose strategies to increase internet trustworthiness and enhance the epistemic value of internet communication.

I begin by defining terminology, delimiting the scope of my argument, and (in section 2) introducing the veritistic social epistemology that provides my evaluative framework. Section 3 outlines veritistic concerns about internet anonymity. In sections 4 and 5, I discuss two leading methods for holding people's real-world selves accountable for the behavior of their online personas. I show that each method has significant epistemic problems. Section 6 concludes by drawing on experimental work in social psychology and behavioral economics to suggest ways that internet accountability could be modified and supplemented to avoid these pitfalls.

1. Preliminary distinctions and scope of the argument

I take anonymity to be a relation of noncoordinatability of traits, such that aspects of an anonymous person's identity are not coordinatable with traits known to others (Wallace 1999: 23). Anonymity thus encompasses pseudonymity, and a user can be anonymous even though she may have a comprehensive and accurate online pseudonymous identity (when the pseudonym prevents others from coordinating the online persona's traits known to them with aspects of the real-world identity, such as the real-world name). While many other media allow for anonymous communication, internet anonymity has particularly epistemically powerful and challenging features. First, unlike anonymous graffiti, anonymous online messages can reach large-scale

audiences, be easily posted from afar, are searchable and reproducible, and are difficult to erase permanently (Levmore 2010: 53). Second, unlike newspaper articles published without an author's name, there is often no intermediary or responsible authority (e.g. an editor) who is held accountable for online speech. Given the ease with which many can post permanent and widely-accessible online messages with impunity, it is unsurprising that many argue for greater accountability for internet speech.

However, this paper argues that accountability mechanisms can damage the epistemic status of internet communities. I focus on two types of internet communities: wiki-style encyclopedia communities and social media communities. Much philosophical discussion of internet epistemology analyzes encyclopedias, such as *Wikipedia*, where users write and edit the articles (cf. de Laat 2012a, 2012b; Magnus 2009; Matthews and Simon 2012; Sanger 2009; Simon 2010; Tollefsen 2009; Wray 2009). My analysis of problems with accountability in these communities challenges common philosophical approaches to internet epistemology. I also analyze social media communities, since blogs, Twitter, and Facebook are growing in both social influence and philosophical interest (cf. de Laat 2008; Goldman 2008; Coady 2011, 2012; Matthews and Simon 2012; Munn 2012; Simon 2010). While *Wikipedia* and social media have many epistemic differences, analyzing the problems of accountability in these different communities uncovers some of the underlying features of internet accountability.

What do I mean by 'accountability mechanisms'? Accountability mechanisms are attempts to increase the trustworthiness of agents grounded in a Hobbesian or economic view of human nature and the solution to its ills. On this *Homo economicus* view, humans are self-interested, rational, utility-maximizing agents. Under certain conditions (e.g., scarcity of resources, when such behavior garners them reputation and other valued goods, etc.), it is in the

self-interest of unfettered rational egoists to engage in antisocial and untrustworthy behavior. But punitive systems create disincentives for such antisocial behaviors. Accordingly, rational choice theorists explain why individuals act trustworthily in certain circumstances by demonstrating that trustworthiness promotes self-interest. A common rational choice strategy shows that trustworthy behavior can be expected in communities where untrustworthy behavior is likely to be detected and punished (Rescher 1989; Blais 1987; Adler 1994; Hardin 2002). Mechanisms to detect and sanction untrustworthy behavior hold members accountable and provide a disincentive for betrayal. Accountability mechanisms in epistemic communities increase detection and punishment mechanisms for betraying epistemic norms (e.g., norms of honesty).

This paper analyzes methods for increasing the trustworthiness of internet communication by holding people's real-world identities accountable for the actions of their online personas. This limits the scope of my discussion in two ways. First, I focus on ways to increase trustworthiness of speakers, rather than ways to increase the discriminatory abilities of readers. Any problem of deception can be approached in two ways: (i) by focusing on the speaker and attempting to make her more honest, or (ii) by focusing on the hearer and attempting to shield her from dishonesty (e.g., by filtering the information she receives) or increase her abilities to detect and reject falsehoods (e.g., by giving her more information about the trustworthiness of speakers).[2] I discuss the former approach here because accountability mechanisms aim to reduce the threat that deceptively false reports,[3] which agents are at risk of

[2] (Simon 2010; Matthews & Simon 2012) address methods for increasing online readers' ability to detect trustworthy communication.

[3] More specifically, I follow the accountability literature by focusing on *intentionally* misleading reports, rather than *accidentally* misleading claims, which can also be veritistically harmful. The accountability literature focuses on this because accountability mechanisms attempt to deter online untrustworthiness by instituting punishment for intentionally misleading reports, thereby changing the incentive structure for agents who intend to choose what is in their self-interest.

believing, will be disseminated. Thus, my discussion concerns communities of agents with imperfect discriminatory abilities in which an increase in false reports is likely to cause an increase in false beliefs.[4] Accountability mechanisms are assessed according to whether they reduce the reporting of false claims. Second, my argument is limited to mechanisms aimed at holding agents' real-world identities accountable for online behavior. Some methods of accountability aim to increase reliability by imposing various sanctions on the online personas themselves (e.g., public shaming of the online persona, or loss of esteem for the online identity) (cf. Brennan and Pettit 2008: 192). But these methods will not be considered here. Many of the problems with holding real-world identities accountable also plague accountability for online identities, and real-world and online identities are often intertwined. But space does not permit me to address the unique problems facing punishment of online personas.

2. The veritistic framework

My argument that accountability mechanisms can be harmful depends on a veritistic, systems-oriented social-epistemic framework. Systems-oriented social epistemology studies the epistemology of epistemic systems, which are social systems that include "social practices, procedures, institutions, and/or patterns of interpersonal influence that affect the epistemic outcomes" of their members (Goldman 2011: 18). Epistemic systems include formal social institutions, such as scientific or legal institutions, as well as informal or amorphous social systems, such as the blogosphere. As a normative enterprise, systems-oriented social epistemology evaluates epistemic systems according to their positive or negative epistemic

[4] In communities in which the frequency of false belief is independent of the prevalence of false reports there is no veritistic reason for accountability mechanisms to reduce the number of false reports. Thus, the paper often works with the assumption that false claims will be believed to some degree, since this is the situation in which accountability mechanisms are proposed as a solution.

outcomes for members. Since it evaluates the outcomes of systems, this epistemology is consequentialist. Assessing the epistemic consequences of various social systems requires a clear specification of the epistemic goods that well-functioning epistemic systems promote. On a veritistic social epistemology, the fundamental epistemic good is true belief.[5] True belief is not only intrinsically valuable, but it is also instrumentally valuable for obtaining effective means to achieve our ends.[6] Accordingly, this paper evaluates the impact of epistemic systems on the formation and dissemination of true beliefs within the community as a whole.

It is outside the scope of this paper to discuss and evaluate all possible versions of veritism (e.g. egalitarian veritism in which the true beliefs of all agents carry equal value vs. elitist veritism in which the true beliefs of some agents are more valuable). In general I wish to be as ecumenical as possible, to show veritists of all stripes the pitfalls of internet accountability. However, two types of veritism must be discussed, because, as I show in sections 4 and 5, they provide different accounts of the epistemic value of diversity within epistemic communities. Since the damage that internet accountability does to diversity will be central to my argument, I outline these two versions of veritism: *error-avoiding veritism* and *truth-seeking veritism* (cf. Coady 2012: 5-7).

This distinction is based on William James' account of "two ways of looking at our duty in the matter of opinion…*We must know the truth*; and *we must avoid error*" (James 2007/1897: 17-18). As James notes, these duties are distinct. For example, some epistemic practices help us reject falsehoods, but they do so by leaving us in ignorance (i.e., suspended judgment), which is distinct from accepting a truth. Such practices fulfill our duty to avoid error, but they violate our

[5] It is beyond the scope of this paper to address objections to the claim that true belief is of fundamental epistemic value rather than alternatives such as justified belief or consensus. See (Goldman 1999, 2002; Coady 2012) for veritistic evaluation of these alternatives.

[6] In addition, true belief is constitutive of certain values, e.g., we value convicting the truly guilty (Goldman 1999: 75).

duty to know the truth. In systems-oriented social epistemology, one may be forced to evaluate the relative epistemic merits of systems that differ only in truth-seeking and error-avoiding merits. For example, consider the relative merits of two hypothetical filtering systems, applied at different times to the same community of agents with imperfect discriminatory abilities. Epistemic system *A*, which has strong filtering mechanisms, might help members avoid false beliefs (by preventing their dissemination), but it filters out so much information (including many true claims) that it leaves members in a position of greater ignorance. System *A* prevents members from forming false beliefs, but at the expense of leaving them with fewer true beliefs and no opinion about many claims. In contrast, epistemic system *B*, which has no filtering mechanisms, exposes members to a greater number of claims (many of them true, but also many of them false). System *B* helps members form many more true beliefs than does system *A*, and it leaves fewer members in a position of suspended judgment about many topics (since they are exposed to some information about the topics), but *B* also lacks the filtering resources to reduce error. If one makes the simplifying assumption that the beliefs under consideration are of equal interest and value along other dimensions, then in evaluating the relative epistemic merits of these two systems, the social epistemologist must determine which is of greater epistemic value: error avoidance or truth attainment.[7]

Error-avoiding veritists prioritize avoiding false belief, while *truth-seeking veritists* prioritize attaining true belief. Distinguishing between error-avoidance and truth-attainment is crucial in recent discussions on internet epistemology. Goldman (2008) argues that the veritistic

[7] To make this hypothetical example a forced choice between filtering systems in which error avoidance competes with truth attainment, it must be the case that strong filtering mechanisms do not make the agents of this community suspicious. Otherwise, it may be the case that when filtering system *A* is applied, the agents erroneously reject as false the true claims that make it through the system, perhaps because they are suspicious of the motives of the filterers. While this qualification is needed to illustrate the forced choice, I take no position on whether this reflects actual agents' responses to stringent filtering systems. I thank an anonymous reviewer for pushing me to clarify the example.

value of traditional media surpasses that of the blogosphere, in part, because the traditional media are superior at filtering out false claims—traditional reporters' practices "reduce the number of errors that might otherwise be reported and believed" (Goldman 2008: 117). David Coady (2011, 2012) counters that Goldman's argument is inadequate to prove the superiority of traditional media. Coady argues that Goldman fails to provide empirical evidence that the blogosphere causes more false beliefs than do traditional media. But Coady also maintains that, "even if it were true that the blogosphere leads people to believe more falsehoods than they otherwise would, it would not follow that its overall impact on them would be harmful, even from a narrowly epistemic point of view" (Coady 2012: 160). Coady reminds us that error-avoidance and truth-attainment are distinct epistemic goals that may come into conflict: "People who confine themselves to a filtered medium may well avoid believing falsehoods (if the filters are working well), but inevitably they will also miss out on valuable knowledge" (Coady 2012: 161). Even if it were true that the blogosphere causes more false beliefs than the traditional media, it might still have the epistemic virtue of causing many true beliefs. Thus, a truth-seeking veritist might remain unconvinced by Goldman's argument for the epistemic inferiority of the blogosphere. Having illustrated the two versions of veritism, I take no side in this debate on the blogosphere, nor the debate between truth-seeking and error-avoiding veritists; instead, sections 4 and 5 show that both types of veritist have reason to be concerned about internet accountability.

3. Veritistic concerns about internet anonymity

Veritists have reason to be concerned that internet anonymity facilitates error dissemination. Anonymity protects agents' real-world identities from sanction for the actions of

their online personas. Tricksters, trolls, vandals, corporations, and special interest groups, emboldened by this protection, can spread falsehoods for fun, harm, self-promotion, spinning information, or political gain. Internet encyclopedias have been the victims of such attempts to spread misinformation (Borland 2007). While the reliability of *Wikipedia* entries rates favorably compared to other encyclopedias such as the *Encyclopedia Britannica* (Giles 2005), it is also vulnerable to vandalism, since anyone can anonymously edit most[8] entries. In one study, it was found that only 4% of all edits to *Wikipedia* pages were cases of vandalism (*Wikipedia* contributors 'Counter-Vandalism'). However, anonymity appears to facilitate vandalism, since 97% of the vandalizing edits were made by anonymous users (ibid.).[9]

Just as Wikipedians can edit pages anonymously, bloggers, Twitterers, and other social media users can participate without providing accurate information about their real-world identities, and users with fake online identities have used this protection to spread misinformation. A well-publicized hoax blog provides an illustration. On February 2, 2011, Amina Abdallah Arraf started a blog titled *A Gay Girl in Damascus*. For four months, Amina wrote about her life as an openly lesbian Syrian-American living in Damascus during the Syrian uprising. Amina's blog was widely read in the West. Western reporters heralded her as a hero of the uprising ('A Gay Girl…' 2011, Bell & Flock 2011). On June 6, a blog post from Amina's cousin reported that Amina had been kidnapped by armed men (Hassan 2011). Several Western news outlets reported Amina's capture, and a campaign began to locate and free her. At one point, a Facebook group calling for her release had 14,000 followers, and the American embassy

[8] Some pages deemed at risk of damage from open editing are placed in 'protected' status, which restricts the kind of edits that can be made and who can edit them (*Wikipedia* contributors 'Protection Policy').
[9] While this study did not determine what proportion of all edits were anonymous edits, an IBM study found that around 31% of the page versions studied were contributed by anonymous unregistered users (Viégas et al. 2004: 580). However, the IBM study noted cases of useful anonymous editing and, without providing statistical support, denied a clear correlation between anonymity and vandalism.

investigated her disappearance (Bell & Flock 2011). But unbeknownst to her supporters, Amina Arraf did not exist. In fact, Tom MacMaster, a white American living in Edinburgh, was the sole author of all posts on the blog.

This internet imposture caused clear epistemic damage. Amina's blog disseminated false claims about the life of someone who never existed. Many readers of the blog believed these claims, and the Western media's promotion of Amina as a hero of the uprising spread many of these false beliefs to a wider audience. But not only did MacMaster's imposture spread false beliefs about Amina, but the imposture also provided an inaccurate picture of the Syrian uprising and LGBT life in Syria. Additionally, the exposure of the hoax provided an opportunity for the Assad regime to perpetuate the conspiracy theory that the Syrian uprising is a foreign intervention in Syrian politics, rather than a citizen revolution (Abbas and Boundaoui 2011). Finally, following exposure of the hoax, the credibility of all uses of social media by Middle Eastern activists was thrown into dispute (Abbas 2011). Thus, the hoax may have undermined trust in many legitimate activists, thus undermining their ability to communicate true claims about the Arab Spring. In all these ways, MacMaster's abuse of the anonymity provided by social media shows that veritists have good reason to fear for the veritistic value of the internet. It is not surprising, therefore, that each time there is a highly publicized internet hoax or case of *Wikipedia* vandalism, there are calls for greater internet accountability as a solution, on the grounds that the threat of punishment for online dishonesty might deter some of these abuses. In sections 4 and 5, I outline two methods of internet accountability, and I argue that each has significant veritistic problems.

4. Accountability mechanism 1: Require identified communication (i.e., abandon anonymity)

Abandoning anonymity is one way to increase internet accountability and protect the trustworthiness of internet communication. This involves replacing anonymous systems with systems that prompt users to interact using real-world identities. While this would be a significant change to many internet communities, a number of existing mechanisms could be used. Several internet wiki encyclopedias require editors to adopt real-name user identities (e.g., *Citizendium*, *Scholarpedia*). Accountability concerns are explicitly listed as one of the reasons for *Citizendium*'s real-names policy: "people do tend to behave themselves better when their identities are known and their behavior is out in the open, and good behavior is crucial to a smoothly running knowledge community" (*Citizendium* contributors 'Why Real Names?'). Similarly, blog moderators can require that commenters sign their comments, and they can delete or refuse to post comments by anonymous commenters. While these strategies can push many community members to non-anonymous participation, the possibility often remains that someone can provide a fake identity, if there are no additional verification mechanisms. To avoid this problem, Twitter has some verified accounts (usually of celebrities and other influential figures) for which the organization has confirmed the user's real-world identity (Twitter 'FAQs').

While abandoning anonymity may increase the reliability of content shared by users afraid to sully their real-world identities, there are epistemic benefits of anonymity that are lost. First, some experimental evidence indicates that anonymity in computer-mediated discussion increases the quantity and novelty of ideas shared (Connolly, Jessup, and Valacich 1990). While the precise reasons for these results are unclear, the authors speculate that all of us feel some degree of social inhibition when our comments are linked to our identities (ibid.: 699). Anonymity enables us to more freely contribute our epistemic resources.

Second, in some areas, particularly vulnerable agents will not transmit their knowledge for fear of punishment, retaliation, or embarrassment. Anonymity has enabled many activists in the Middle East to engage in citizen reporting and share information with other activists (Howard and Hussein 2011). Similarly, citizen reporters have used Twitter to disseminate information about drug-related violence in Mexico (Mustafaraj et al. 2012). Reporters on drug cartels in Mexico have been targeted for attacks, including citizen journalists who have been murdered by those attempting to silence social networking communities (Associated Press 2011). In such threatening circumstances, anonymity provides a measure of safety. Without it, knowledge may not be shared. Consider the case of HarassMap, a social initiative to collect and share information about incidents of street harassment in Egypt (HarassMap.org). Victims of street harassment can report incidents by using the HarassMap website, or sending an SMS text, Tweet, or e-mail. The platform then aggregates this information and displays locations of types of harassment visually on a map. The site provides crucial information about the extent and nature of street harassment. Given the social environment in which women routinely have their claims of harassment dismissed, ignored, or labeled as their own fault, the anonymity provided by HarassMap is central to its use as a knowledge dissemination mechanism. Additionally, internet anonymity has provided opportunities for LGBT people to build community and share knowledge (cf. Whittle 1998). The discrimination and violence to which LGBT individuals are often subjected makes anonymity an important protection. These are all examples of knowledge bearers who have legitimate concerns of serious sanction were their testimony connected with their real-world identities. Note that, in these four examples, those who are protected by anonymity lack some form of social power, or belong to a marginalized or oppressed group.

This provides an argument from truth-seeking veritism against accountability mechanisms that remove anonymity. Removal of anonymity could deprive the community of true beliefs spread by reports from socially threatened groups. Without online anonymity, activists, citizen journalists, and members of many socially stigmatized groups are much less likely to take the risk of sharing what they know with others. Thus, the decrease in threatened groups' participation in epistemic communities is a second problem with removing anonymity. Abandoning anonymity can decrease the diversity of epistemic communities.

At this point, one might wonder whether my argument will only be persuasive to truth-seeking veritists. An objector might claim that I am rejecting the main mechanism for preventing people from adopting false beliefs (i.e., preventing the dissemination of falsehoods by the deterrent of accountability mechanisms) in order to maintain a system that allows us to form more true beliefs. But what about error-avoiding veritism? If forced to choose between an epistemic system that prevents agents from believing falsehoods but leaves them in a position of suspended judgment (ignorance) on many subjects, and an epistemic system that facilitates the formation of beliefs but does not weed out falsehoods, error-avoiding veritists will choose the former.[10] If the decision whether to remove anonymity is such a forced choice, then error-avoiding veritists will *ceteris paribus* prefer the accountability mechanism, even though it has the pitfall of preventing the dissemination of true beliefs. Thus, to show that error-avoiding veritists also have reason to be wary of the removal of anonymity, I argue that this forced choice is rare. In other words, anonymity facilitates error detection as well as truth attainment.

[10] More precisely, which system error-avoiding veritists will accord more epistemic value depends on both the weight to which they assign true and false beliefs as well as the number of true and false beliefs promoted by the systems. Only the strictest error-avoiding epistemologists will place all epistemic value on error-avoidance and accord no weight to the adoption of true beliefs. Thus, error-avoiding veritists will vary based on the weights they assign to truth acquisition and error-avoidance. However, what unites them is that there exists some number of false beliefs promoted by an epistemic system that cannot be outweighed by the value of removing ignorance.

Accordingly, a third problem with requiring identified communication is that anonymity can enhance error-detection by enabling increased transformative criticism to weed out error and bias. Feminist empiricist epistemologists have long argued that critical dialogue within diverse epistemic communities facilitates the uncovering and removal of bias (e.g., Longino 1990, 2002; Anderson 1995; Intemann 2010). In scientific communities, diversity is epistemically valuable because biased background assumptions can be uncovered when research is subjected to conceptual and evidential criticism from scientists with diverse values, interests, and social identities. If the scientific community promotes such critical engagement within a diverse community, and if criticisms of bias are taken up, then the knowledge circulating in the community can become more objective (Longino 1990: 73-74). As Goldman notes, this feminist argument can be given a veritistic interpretation (Goldman 1999: 78)—critical discussion within diverse epistemic communities is an effective mechanism for error detection. Insofar as online anonymity promotes diversity within online communities, it can facilitate error detection. In contrast to the truth-seeking veritistic argument that anonymity has epistemic merit because it encourages threatened groups to share *true reports* that can spread true beliefs to others, the error-avoiding veritistic argument maintains that anonymity enables such groups to share *criticisms* of false beliefs. These criticisms can lead community members to reject or suspend judgment on false claims.

Blogging and tweeting are not simply means of disseminating knowledge claims; they are also means of challenging, criticizing, and uncovering errors in others' knowledge claims. Similarly, Wikipedians do not simply post and edit entries, they use the 'Talk' page of encyclopedia entries to challenge content and debate bias in the articles. Thus social media and *Wikipedia* provide ample opportunities for transformative criticism. The error-uncovering

efficacy of such criticism is enhanced by the anonymity that facilitates participation by diverse groups who would otherwise, for fear of sanction, not join the discussion. Removing anonymity risks silencing their valuable criticisms.

The previous three arguments defending internet anonymity focused on anonymity's role in enabling threatened groups to share true reports and criticisms. A fourth argument emphasizes anonymity's role in ensuring that these reports and criticisms are given due epistemic authority. As recent work on epistemic injustice has shown, prejudice causes socially oppressed groups to be given less epistemic authority than they deserve (Fricker 2007). This is veritistically undesirable, since such groups make valuable reports and criticisms that merit uptake.

Consider testimonial injustice: a speaker sustains a testimonial injustice iff "she receives a credibility deficit owing to identity prejudice in the hearer" (Fricker 2007: 28). When a hearer's prejudicial stereotypes cause her to grant the speaker's testimony less credibility than would have been granted in the absence of prejudice, the speaker suffers an injustice because she is undermined in her capacity as a knower (Fricker 2007: 44; 2011: 67). The community also suffers, since the knowledge the speaker attempted to convey does not receive appropriate uptake (Fricker 2007: 43). Recent work on implicit bias shows that prejudice is not only reflected in a hearer's explicitly held beliefs, but can also operate unconsciously.[11] Blinding policies can reduce testimonial injustice: when information about the speaker's identity is withheld from the hearer, it is less likely[12] for the hearer's stereotypes to be engaged against the speaker. Internet anonymity facilitates such blinding. Members of groups that commonly experience testimonial injustice can blog, tweet, and edit *Wikipedia* pages using identities that do not trigger stereotypes

[11] See (Kelly and Roeddert 2008) and (Project Implicit n.d.) for useful overviews of the concept of implicit bias and related literature.

[12] The hearer may still, consciously or unconsciously, pick up on cues in the speaker's speech that lead the hearer to draw conclusions about the identity of the speaker and thus trigger prejudicial stereotypes.

that would undermine their credibility (cf. Wallace 1999: 28; Klein, Clark & Herskovitz 2003: 358; Matthews & Simon 2012:48). For example, a woman who may find her reports or criticisms dismissed as emotional or less competent in non-anonymous contexts can post anonymously and have her claims given due uptake.

In sum, my initial arguments showed that abandoning anonymity can undermine the sharing of true beliefs and the detection of false beliefs by decreasing the epistemic activity of threatened epistemic agents. While those arguments were based on the premise that such groups will participate less in non-anonymous epistemic communities, this fourth argument maintains that their participation will, without anonymity, be given less epistemic authority than it deserves. Anonymity, thus, has veritistic benefits when it facilitates the dissemination and uptake of knowledge that would otherwise be lost to the community.

But what of the bad behavior facilitated by anonymity? One might argue that anonymity provides no net benefit through the mechanisms I have described, because while it can encourage marginalized groups to participate in epistemic communities and have their claims taken seriously, it also simultaneously encourages attacks on them. Bad faith actors, emboldened by the safety of anonymity, can spread prejudice and hate speech, making online communities hostile environments for oppressed groups.[13] Hate speech and harassment of oppressed groups can cause them to leave epistemic communities or to stay but internalize the attacks and lose confidence in their contributions (Anderson 1995: 201). It also spreads prejudice, further entrenching problems of testimonial injustice. Thus, anonymity may provide no net increase in the participation of oppressed groups in internet communities and no net increase in their

[13] For discussion of harassment in anonymous computer-mediated communication, see (Wallace 1999: 32; Klein, Clark & Herskovitz 2003: 373; Citron 2010).

epistemic activity being given due credit. Thus, on this objection, there is no clear veritistic argument, along the lines I have suggested, in favor of internet anonymity.

In response, I grant that, in the absence of norms of civility, anonymity may have no net epistemic benefit. However, this should be taken as an argument for internet norms of civility, rather than an argument against anonymity. When internet communities are set up to prevent hate speech and harassment of oppressed groups, the problems outlined in the objection are not as pressing. A variety of mechanisms for preventing hate speech and harassment exist, including the clear articulation of norms of civil dialogue and the institution of reporting mechanisms so that harassing speech is flagged for removal by a moderator.[14] This is not to say that ensuring a welcoming internet community for marginalized groups is easy, but steps can be taken to diminish internet harassment without removing the anonymity that provides protection to vulnerable groups.[15]

Given the veritistic problems with requiring identified communication, it is perhaps fortunate that *Wikipedia* and many social media spaces have not completely removed anonymity. That said, other mechanisms exist for holding people's real-world selves accountable for the activity of their online personas. The next section provides a veritistic analysis of another prominent mechanism: investigative accountability.

5. Accountability mechanism 2: Investigative accountability

[14] Extreme, coordinated harassment by groups of internet attackers may be facilitated by anonymity and impervious to such anti-harassment measures (Citron 2010; Sarkeesian 2012). If coordinated harassment becomes an overwhelming problem for oppressed groups, then anonymity may become a net epistemic detriment.

[15] One might wonder whether these solutions to internet harassment simply smuggle internet accountability in through the back door. Often the articulation of social norms is the first part of accountability mechanisms, as they lay out standards to which agents are held accountable. In addition, removing hate speech is one way to punish an agent for violating these standards. In response, I issue a promissory note that section 6 argues that some accountability mechanisms can be valuable if balanced with other ways of inspiring trustworthy behavior. I also provide another interpretation of codes of ethics and social norms according to which they have value beyond their function in accountability mechanisms.

While it is common to describe the internet as affording anonymity to users, it is more accurate to describe much internet activity as only potentially anonymous. This is due to the proliferation of increasingly sophisticated methods for uncovering the real-world identities behind online personas. Such investigation usually requires more effort (and expertise) than most of the general public are willing (or able) to provide. Moreover, the very existence of detection methods is often not widely known. In many cases, although someone might believe that their internet activity is anonymous, it is only anonymous as long as those with the resources and ability to investigate their real-world identity are not moved to do so.

For example, WikiScanner is a tool that enables users to identify the source of some *Wikipedia* edits (Griffith 2008; Simon 2010). *Wikipedia* allows people to edit pages anonymously by tagging their edits with only the IP address from which the edit was made. WikiScanner automatically combines the database of IP addresses of *Wikipedia* edits with the database of the companies that own the IP addresses. Thus this tool is a boon to investigators aiming to discover who is behind edits to *Wikipedia* pages. A related tool by the makers of WikiScanner, the Poor Man's CheckUser, reveals the IP addresses behind some registered Wikipedians' accounts (Griffith n.d.). This tool also facilitates investigation into the real-world identities behind *Wikipedia* activity.

The investigation into the fake *A Gay Girl in Damascus* blog illustrates other means of investigative accountability. Not everyone was fooled by MacMaster's imposture. Soon after Amina's capture by Syrian authorities was reported, Daniel Nassar,[16] a gay Syrian, began to have doubts about her existence (Nassar 2011). He asked around in the Syrian lesbian community and found no one who knew Amina, which raised suspicions since Amina presented herself as active in these circles. Nassar contacted NPR reporter Andy Carvin with his

[16] 'Daniel Nassar' is a pseudonym.

suspicions, and Carvin tweeted a query asking whether anyone had met Amina in person or talked to her on the phone. When no confirmation of her identity followed, several blogs and news organizations began searching for Amina. The investigation was intense. For example, the post on Liz Henry's blog (which became a site of crowd-sourced investigation) has over 400 related comments (Henry 2011a). Investigators uncovered property records, IP addresses, photos, and Facebook posts linking Tom MacMaster to Amina. Finally, in the face of this mounting evidence, MacMaster admitted to the hoax.

In response to cases of *Wikipedia* vandalism and fake blogs, some have advocated wider use of both WikiScanner and the verification of bloggers' real-world identities (de Laat 2012a; Simon 2010; Henry 2011b; York 2011). They argue that suspicious *Wikipedia* edits and blog posts should be investigated to determine the source's identity. On this line of argument, if agents know that suspicious activity on their part will trigger the uncovering of their identity, then they will be less likely to spread falsehoods and engage in other epistemically harmful acts. The second method of accountability is, thus, to ensure trustworthiness through the threat of having one's real-world identity uncovered by investigation into suspicious online behavior. This kind of accountability has some veritistic merit insofar as fear of exposure may inhibit some dishonest online activity—but are there any problems with it?

First, investigations are distracting, squandering scarce resources (cf. O'Neill 2002: 50). This accountability approach is premised on a well-functioning detection mechanism that will alert the relevant punitive agents (e.g., moderators or the public) to the existence of a bad faith actor. But detection is time and resource intensive. In the Amina case, a small army of bloggers, journalists, and activists was caught up in the hunt to uncover the blogger's identity. The six-day-long investigation ate up many hours of people's lives and much mental energy. Some of

those involved recount being exhausted and losing sleep during the investigation (Henry 2011a). Some expressed frustration that so much attention was diverted from investigation into the real events of the Syrian uprising (Henry 2011b). Thus, the investigative approach requires that the community sink significant time into detection practices, which means that time and resources will be diverted from other worthy epistemic aims.[17] This argument can be made by both truth-seeking veritists and error-avoiding veritists—investigation diverts epistemic resources, thereby preventing the community from obtaining other valuable truths and detecting other important errors.[18]

Second, investigative accountability can have a dampening effect on internet speech as those who desire anonymity avoid making surprising claims that might raise the suspicions of potential investigators. Consider two of the ways we evaluate the reliability of internet content. First, we assess internet claims by the plausibility of their content (Magnus 2009: 81). If, in light of our background beliefs, a claim seems implausible, then we are likely to suspect it is the product of an incompetent or bad faith actor. In addition, we engage in sampling—claims that are repeated by other independent sources are regarded as more plausible than claims made by lone voices (ibid.: 82). Thus, the claims likely to raise suspicions, and thereby trigger investigations, are those that are surprising and novel. In an environment of widespread investigative accountability, the best strategy to protect one's anonymity is to avoid making novel and surprising claims. This is veritistically harmful, since novel and surprising truths are interesting in their own right (thus of particular value to truth-seeking veritists) and also

[17] Note that as investigative accountability becomes easier (thereby decreasing the epistemic burden), we face the problems attendant to removing anonymity, as previously described (section 4).
[18] In addition, it shows that this method of accountability potentially lacks the merits of speed and efficiency, which are important standards of social epistemic appraisal (Goldman 1992: 195-96).

suggestive of transformative criticisms that uncover previously unrecognized errors (thereby of particular value to error-avoiding veritists) (cf. Goldman 1999: 107).

The speech of socially threatened groups is particularly susceptible to this dampening effect. Often those who most need the protection of anonymity are precisely those whose claims are most likely to be seen as novel and surprising, and hence suspicious. Anonymity protects marginalized and stigmatized minorities from retaliation and testimonial injustice. But their accounts of their lives (and their criticisms of dominant viewpoints) are less likely to conform to the background assumptions of mainstream readers. Thus, their claims are likely to fail many readers' tests of plausibility of content. Moreover, as underrepresented and marginalized groups, they are likely to make claims that are not widely repeated. When a reader samples other internet sources to corroborate the claims, they are more likely to come up empty-handed. Thus, the claims of marginalized and stigmatized groups are likely to also fail many readers' sampling tests.

The investigation into *A Gay Girl in Damascus* provides illustration of the suspicion cast on marginalized groups. As Liz Henry (one of the investigators) reports, some of the initial suspicions about the blog's authenticity cited doubts that there could be an out lesbian in Damascus (Henry 2011a). In light of persistent attempts to render the LGBT community in the Middle East invisible by denying their existence,[19] this reason for doubting Amina's blog is troubling. Additionally, one investigator raised doubts about the existence of Amina's cousin because the cousin, a married woman, belonged to lesbian Facebook groups (Henry 2011a). As Henry rightly objected at the time, this reason for doubt ignores the real existence of married bisexual women. Thus, this investigation reveals a pattern of doubting the authenticity of

[19] One prominent example is President Ahmadinejad's denial of homosexuality in Iran (Fathi 2007).

stigmatized minorities. While the speech of straight Syrian activists went uninvestigated, the speech of LGBT Syrians was viewed as suspicious.

The case also provides anecdotal evidence that vulnerable members of online communities withdraw from participation when the threat of investigation increases. Vulnerable users often attempt to protect themselves by 'flying under the radar,' e.g., targeting their blogs to a specific audience and avoiding drawing broader attention. For example, LGBT Syrian bloggers risk discrimination if their LGBT identity is exposed in the real world, and political blogging can draw harassment by authorities. Bloggers protect themselves by using pseudonyms and avoiding political posts that might trigger investigation (Hamwi and Nassar 2011). But once Amina's blog gained international attention, LGBT bloggers came under the spotlight. In response, Daniel Nassar, the gay Syrian blogger who raised doubts about Amina's authenticity, changed his online behavior to protect himself. He reports going "back to the closet on all the social media websites I use" (Hamwi and Nassar 2011). Thus, drawing additional attention to the internet speech of vulnerable users can have negative veritistic consequences—it can prompt them to censor their speech. So investigations are not always epistemically innocent.

One might object that this is a poor case study on which to base my argument, since Amina and her cousin were in fact fictional. So perhaps the suspicion cast on these LGBT online identities was epistemically productive in uncovering falsehoods. But this objection misses the thrust of my argument. I do not deny that in some cases there may be epistemic benefits from investigative accountability. My aim is to show that it can also have harmful epistemic effects. It can decrease the diversity of internet speech, thereby undermining both truth attainment and error detection. Additional empirical research is needed, and careful analysis is required before we advocate accountability solutions to the problem of internet untrustworthiness. For example,

while some false beliefs were corrected by exposure of the Amina hoax, how many more true beliefs are now less likely to be produced, disseminated, and taken up as vulnerable agents withdraw from online communities for fear of exposure? Would widespread investigative accountability of bloggers and Wikipedians do more harm than good? These questions must be answered, once we recognize the possible pitfalls of internet accountability.

Where does this leave us with internet trustworthiness? I have shown that the two leading accountability mechanisms for online communities have serious epistemic problems. While I have not shown that every conceivable accountability mechanism is epistemically damaging, I have shown that the most prominent exemplars of this approach have risks. In addition, other accountability mechanisms are likely to share some of these problems, since holding users' real-world selves accountable will generally require uncovering their real-world identity. Thus, accountability mechanisms should be carefully evaluated before implementation to determine whether their benefits (e.g., deterring untrustworthiness) outweigh their harms (e.g., distraction and less diverse online communities).

But can anything more than this be done? Can accountability mechanisms be modified to avoid their potential harms? Are there alternatives to accountability mechanisms? In the final section of the paper, I argue that not only are some modifications available, but, if we turn to behavioral economics and social psychology, we can generate a different approach to internet trustworthiness. By looking to ways to engage users' pro-social motivations, we can add another set of tools to our social-epistemic toolbox. I suggest some ways to modify and balance investigative accountability with other approaches.

6. Towards modified and balanced accountability

Investigative accountability could play a positive epistemic role in online communities, but only if other measures are taken to protect veritistically valuable diversity. To prevent the problem of disproportionate investigation of marginalized and minority users, epistemic communities need mechanisms for checking the biases of potential investigators. While some spurious investigations could be prevented by potential investigators cultivating epistemically responsible habits of checking their own biases, there are limits to the ability of individual agents to monitor their own prejudices. Thus, more communal solutions to the problem of bias are required. For example, if the question of whether some internet speech merits investigation is debated within a community, then as the diversity of that community increases, the likelihood increases that biased reasons for suspicion will be challenged. Recall that internet speech by minorities is likely to fail the plausibility and sampling tests of mainstream users. But those with non-mainstream background assumptions and greater familiarity with internet subcultures are likely better able to show that the speech can pass such tests. Liz Henry's criticism of others' suspicion of LGBT speech illustrates this. By questioning the assumptions that there are no lesbians in Syria or that married women do not belong to internet lesbian groups, she provided a useful check on biased investigative accountability. I, therefore, suggest that decisions to investigate online identities be made in the context of debate within a diverse community.

But what of the problem of distraction? Investigations divert time and resources from other epistemic projects. My proposal of adding critical dialogue to investigative accountability adds to the epistemic burden. If investigation is our only tool to increase internet trustworthiness, then significant resources must be devoted to it to maintain a strong threat of having one's real-world identity revealed.[20] Given the sheer volume of internet discourse, the

[20] One might argue that we could increase trustworthiness simply by increasing the punishment for untrustworthiness, and that this might distract fewer resources than increased investigation. However, research

demands are daunting. However, less investigation is needed if we balance accountability mechanisms with other ways of promoting internet trustworthiness.

So what are the alternatives? Looking beyond the rational choice model reveals other options. Recent work in behavioral and experimental economics, sociology, and psychology has challenged the *Homo economicus* model of human behavior. There is now widespread empirical evidence that humans are not driven by self-interest alone; we are also motivated by altruistic and pro-social concerns (cf. Batson 2002; Gneezy 2005; Ostrom 2005; Mazar and Ariely 2006; Mazar, Amir, and Ariely 2008). This intrinsic concern for others is not always operative, and it can be overwhelmed by self-interested concerns. However, experimental work has uncovered some ways to engage people's pro-social motivations, even when there are temptations for untrustworthiness and dishonesty. This research suggests two ways to improve internet trustworthiness by engaging pro-social motivations.

First, there is experimental evidence that "people care not only about how much they gain from a lie, but also how much the other side loses" (Gneezy 2005; 391). People appear less likely to lie when the lie only gives them a small benefit but does the recipient a great harm (ibid.: 385). The internet is, in many ways, an impersonal mode of interaction—users communicate with an audience that is often completely unknown to them. This may make it difficult for people to be cognizant of the damage internet dishonesty does. The experimental results suggest that making the significant harms of internet untrustworthiness more salient to users might activate the pro-social motivation to avoid doing damage to others. Of course, if a user believes that she has much to gain from dishonesty, then this particular psychological mechanism will not overcome the temptation. But I hypothesize that many cases of internet

---

shows that of the two components of accountability mechanisms—detection and punishment—the probability of detection has a stronger deterrent effect (Nagin and Pogarsky 2003).

dishonesty do not provide large benefits. Many users spread misinformation as a simple joke, prank, or curious attempt to see what would happen. To avoid such low-benefit dishonesty, it might help to display faces and stories of people who have been hurt by internet dishonesty (e.g., students misled by *Wikipedia* vandalism or Syrian activists hurt by MacMaster's hoax). If users encounter such reminders in prominent places (e.g., on banners or pop-up windows that appear when a user begins to edit a page), their desire not to hurt others when it benefits them little may discourage dishonesty. Research shows that people are more altruistic when shown a photograph of the person whom their action affects—this suggests that making potential victims visible may help (Burnham 2003).[21]

Second, experimental research shows that when people are asked to reflect on their own moral values or read a code of ethics before being tempted with an opportunity for profitable deception, they are less likely to be dishonest, even when there is no risk of dishonesty being detected (Mazar and Ariely 2006; Mazar, Amir, and Ariely 2008). Such research is relevant for internet trustworthiness—mechanisms could be added to increase trustworthiness even when people do not fear detection. Such mechanisms could remind users of their own values or lead them to read some ethical principles of trustworthy internet behavior before posting material. While many online communities have behavioral policies, guidelines, or visions for the kind of community they want to sustain, these ethical codes usually remain in the background of communication. Users can access them by searching for the policy statement pages of blogs, *Wikipedia*, or Twitter, etc. (e.g., BlogHer 'What are Your…'; *Wikipedia* Contributors 'Category: Wikipedia Behavioral'; Twitter 'The Twitter Rules'). But one rarely encounters these codes of ethics unless one is being chastised or punished for violation. They function largely as standards

[21] Burnham posits an ultimately self-interested evolutionary story behind this altruism (Burnham 2003:141). It is outside the scope of this paper to address the question of whether our altruistic psychology ultimately has egoistic roots (cf. Sober and Wilson 1998; Batson 2002).

to which agents are held accountable, or as policies to protect the website owners from liability. I suggest another possible use for such policies: reminders of values that users themselves deem worthy of respect—tools for triggering users' pro-social motivations. Online communities could bring ethical prohibitions on dishonesty out from the shadows of the "about us" pages and draw users' attention to them when temptation for dishonesty arises. Such reminders of users' moral values might decrease internet dishonesty.

In conclusion, social epistemologists need ways to ensure internet trustworthiness other than accountability mechanisms. Accountability mechanisms attempt to promote trustworthiness by holding people's real-world identities liable for the behavior of their online personas. But accountability mechanisms have significant epistemic problems. Anonymity provides protection for many groups who have much of epistemic value to contribute. Removing or considerably decreasing this protection through accountability mechanisms threatens the diversity of online communities, diversity which enables communities to both obtain true beliefs and weed out errors. Investigations into real-world identities can waste epistemic resources and be biased against marginalized groups. If investigations are to be part of an epistemically fruitful community, they must be coupled with mechanisms to avoid bias. But also, we will not need to implement widespread, potentially damaging accountability mechanisms if we also pursue other mechanisms to engage users' pro-social motivations. Humans are not simply self-interested agents who can only be spurred to trustworthiness through threat of punishment. Research on the triggers for users' moral impulses suggests ways to increase the veritistic value of the internet, while avoiding the damage of excessive accountability.

References

‘‘A Gay Girl in Damascus’ How the Hoax Unfolded.’ 2011. *The Telegraph*. Retrieved June 27, 2011, from http://www.telegraph.co.uk/news/worldnews/middleeast/syria/8572884/A-Gay-Girl-in-Damascus-how-the-hoax-unfolded.html.

Abbas, F. 2011. ‘Let Us Not Allow ‘A Gay Girl in Damascus’ to Discredit All Blogging.’ *The Huffington Post*. Retrieved June 27, 2011, from http://www.huffingtonpost.com/faisal-abbas/let-us-not-allow-a-gay-gi_b_880215.html.

Abbas, A. and Boundaoui, A. 2011. ‘A Gay (Straight) Girl (Man) in Damascus (Edinburgh): The Politics behind the Roleplay.’ *KABOBfest*. Retrieved March 4, 2012, from http://www.kabobfest.com/2011/06/a-gay-girl-in-damascus.html.

Adler, J. 1994. ‘Testimony, Trust, Knowing.’ *Journal of Philosophy*, 91 (5): 264-275.

Anderson, E. 1995. ‘The Democratic University: The Role of Justice in the Production of Knowledge, Social Philosophy and Policy.’ *Social Philosophy & Policy*, 12 (2): 186-219.

Associated Press. 2011. ‘Woman’s Decapitation Linked to Web Posts about Mexican Drug Cartel.’ *The Guardian*. Retrieved January 20, 2013, from http://www.guardian.co.uk/world/2011/sep/25/mexico-woman-decapitated-social-network.

Batson, C.D. 2002. ‘Addressing the Altruism Question Experimentally.’ In S.G. Post, L.G. Underwood, J.P. Schloss, and W.B. Hurlbut (eds), *Altruism and Altruistic Love: Science, Philosophy, and Religion in Dialogue*. New York: Oxford University Press.

Bell, M. and Flock, E. 2011. ‘‘A Gay Girl in Damascus’ Comes Clean.’ *Washington Post*. Retrieved July 24, 2011, from http://www.washingtonpost.com/lifestyle/style/a-gay-girl-in-damascus-comes-clean/2011/06/12/AGkyH0RH_story_1.html.

Blais, M. 1987. ‘Epistemic Tit for Tat.’ *Journal of Philosophy*, 84 (7): 363-375.

BlogHer n.d. ‘What are Your Community Guidelines?’ Retrieved December 20, 2012, from http://www.blogher.com/what-are-your-community-guidelines.

Borland, J. 2007. ‘See Who’s Editing *Wikipedia* – Diebold, the CIA, a Campaign,’ *Wired*. Retrieved January 24, 2012, from http://www.wired.com/politics/onlinerights/news/2007/08/wiki_tracker?currentPage=all.

Brennan, G. and Pettit, P. 2008. ‘Esteem, Identifiability, and the Internet.’ In J. van den Hoven and J. Weckert (eds), *Information Technology and Moral Philosophy*. New York: Cambridge University Press.

Burnham, T. C. 2003. ‘Engineering Altruism: A Theoretical and Experimental Investigation of Anonymity and Gift Giving.’ *Journal of Economic Behavior & Organization*, 50(1): 133-144.

***Citizendium* contributors**. ‘Why Real Names?’ *Citizendium.* Retrieved December 20, 2012, from http://en.citizendium.org/wiki/CZ:FAQ#Why_real_names.3F.

Citron, D.K. 2010. 'Civil Rights in Our Information Age.' In S. Levmore and M. C. Nussbaum (eds), *The Offensive Internet: Speech, Privacy, and Reputation*. Cambridge, MA: Harvard University Press.

Coady, D. 2012. *What to Believe Now: Applying Epistemology to Contemporary Issues*. Malden, MA: Wiley-Blackwell.

———. 2011. 'An Epistemic Defence of the Blogosphere.' *Journal of Applied Philosophy*, 28 (3): 277-294.

Connolly, T., Jessup, L. M., and Valacich, J.S. 1990. 'Effects of Anonymity and Evaluative Tone on Idea Generation in Computer-Mediated Groups.' *Management Science*, 36 (6): 689-703.

de Laat, P. B. 2008. 'Online Diaries: Reflections on Trust, Privacy, and Exhibitionism.' *Ethics and Information Technology*, 10 (1): 57-69.

———. 2012a. 'Open Source Production of Encyclopedias: Editorial Policies at the Intersection of Organization and Epistemological Trust.' *Social Epistemology*, 26 (1): 71-103.

———. 2012b. 'Coercion or Empowerment? Moderation of Content in *Wikipedia* as 'Essentially Contested' Bureaucratic Rules.' *Ethics and Information Technology*, 14 (2): 123-135.

Fathi, N. 2007. 'Despite Denials, Gays Insist They Exist, if Quietly, in Iran.' *New York Times*. Retrieved January 19, 2013, from http://www.nytimes.com/2007/09/30/world/middleeast/30gays.html.

Fricker, M. 2007. *Epistemic Injustice: Power and Ethics in Knowing*. Oxford: Oxford University Press.

Giles, J. 2005. 'Internet Encyclopaedias Go Head to Head.' *Nature*, 438 (7070): 900–1.

Gneezy, U. 2005. 'The Role of Consequences.' *The American Economic Review*, 95 (1): 384-394.

Goldman, A. I. 1999. *Knowledge in a Social World*. New York: Oxford University Press.

———. 2002. *Pathways to Knowledge*. New York: Oxford University Press.

———. 2008. 'The Social Epistemology of Blogging.' In J. van den Hoven and J. Weckert (eds), *Information Technology and Moral Philosophy*. New York: Cambridge University Press.

———. 2011. 'A Guide to Social Epistemology.' In A. I. Goldman and D. Whitcomb (eds), *Social Epistemology: Essential Readings*. New York: Oxford University Press.

Griffith, V. 2008. 'WikiScanner FAQ.' Retrieved December 7, 2012, from http://virgil.gr/31.

———. n.d. 'WikiWatcher FAQ.' Retrieved December 7, 2012, from http://virgil.gr/67.html.

Hamwi, S. and Nassar, D. 2011. "From Damascus with Love: Blogging in a Totalitarian State.' *Gay Middle East*. Retrieved March 3, 2012, from http://gaymiddleeast.com/news/news%20317.htm.

HarassMap.org n.d. 'HarassMap Executive Summary.' Retrieved December 20, 2012, from http://harassmap.files.wordpress.com/2008/12/harassmap_executive_summary.pdf.

Hardin, R. 2002. *Trust and Trustworthiness*. New York: Russell Sage Foundation.

Hassan, N. 2011. 'Syrian Blogger Amina Abdallah Kidnapped by Armed Men.' *The Guardian.* Retrieved April 3, 2012, from http://www.guardian.co.uk/world/2011/jun/07/syrian-blogger-amina-abdallah-kidnapped.

Henry, L. 2011a. 'Painful Doubts about Amina.' Retrieved March 3, 2012, from http://bookmaniac.org/painful-doubts-about-amina/.

———. 2011b. 'Gay Girl in Damascus Blogging Hoax: Chasing Amina.' Retrieved March 4, 2012, from http://www.blogher.com/gay-girl-damascus-blogging-hoax-chasing-amina.

Howard, P. and Hussein, M. 2011. 'The Role of Digital Media.' *Journal of Democracy,* 22 (3): 35-48.

Intemann, K. 2010. '25 Years of Feminist Empiricism and Standpoint Theory: Where Are We Now?' *Hypatia* 25 (4):778-796.

James, W. 2007/1897. *The Will to Believe and Other Essays in Popular Philosophy*. New York: Cosimo.

Kaczynski, A. 2012. 'Councilman Pushes for Charges against Twitter User Who Spread Falsehoods.' *Buzzfeed.com*. Retrieved January 28, 2012, from http://www.buzzfeed.com/andrewkaczynski/councilman-pushes-for-charges-against-twitter-user.

Keen, A. 2012. 'Twitterers: Take Responsibility for Your Reckless Claims.' *CNN.com*. Retrieved February 1, 2013, from http://www.cnn.com/2012/11/27/opinion/twitter-war-keen/index.html.

Kelly, D. and Roeddert, E. 2008. 'Racial Cognition and the Ethics of Implicit Bias.' *Philosophy Compass*, 3(3): 522-540.

Klein, E., Clark, C., and Herskovitz, P. 2003. 'Philosophical Dimensions of Anonymity in Group Support Systems: Ethical Implications of Social Psychological Consequences.' *Computers in Human Behavior*, 19: 355-382.

Levmore, S. 2010. 'The Internet's Anonymity Problem,' In S. Levmore and M. C. Nussbaum (eds), *The Offensive Internet: Speech, Privacy, and Reputation*. Cambridge, MA: Harvard University Press.

Longino, H. 1990. *Science as Social Knowledge*. Princeton, NJ: Princeton University Press.

———. 2002. *The Fate of Knowledge.* Princeton, NJ: Princeton University Press.

Magnus, P.D. 2009. 'On Trusting *Wikipedia*.' *Episteme*, 6(1): 74-90.

Matthews, P. and Simon, J. 2012. 'Evaluating and Enriching Online Knowledge Exchange: A Socio-epistemological Perspective.' In A. Lazakidou (ed), *Virtual Communities, Social Networks and Collaboration*. New York: Springer.

Mazar, N. and Ariely, D. 2006. 'Dishonesty in Everyday Life and Its Policy Implications.' *Journal of Public Policy & Marketing*, 25 (1): 117-126.

Mazar, N., Amir, O., and Ariely, D. 2008. 'The Dishonesty of Honest People: A Theory of Self-Concept Maintenance.' *Journal of Marketing Research*, 45(6): 633-644.

Munn, N. J. 2012. 'The New Political Blogosphere.' *Social Epistemology*, 26 (1): 55-70.

Mustafaraj, E., Metaxas, P., Finn, S., and Monroy-Hernández, A. 2012. 'Hiding in Plain Sight: A Tale of Trust and Mistrust inside a Community of Citizen Reporters.' *Sixth International AAAI Conference on Weblogs and Social Media*: 250-257.

Nagin, D. S. and Pogarsky, G. 2003. 'An Experimental Investigation of Deterrence: Cheating, Self-Serving Bias, and Impulsivity.' *Criminology*, 41 (1): 501-27.

Nassar, D. 2011. 'Foreign Policy: Damascus Still Has Gay Girls.' *NPR.org*. Retrieved March 30, 2012, from http://www.npr.org/2011/06/16/137217280/foreign-policy-damascus-still-has-gay-girls.

O'Neill, O. 2002. *A Question of Trust*. New York: Cambridge University Press.

Ostrom, E. 2005. 'Policies that Crowd out Reciprocity and Collective Action.' In H. Gintis, S. Bowles, R. Boyd, and E. Fehr (eds), *Moral Sentiments and Material Interests*. Cambridge, MA: MIT Press.

Project Implicit. n.d.. Project Implicit Information Website. Retrieved January 23, 2012, from http://www.projectimplicit.net/index.php.

Rescher, N. 1989. *Cognitive Economy: An Inquiry into the Economic Dimension of the Theory of Knowledge.* Pittsburgh, PA: University of Pittsburgh Press.

Sanger, L. M. 2009. 'The Fate of Expertise after *Wikipedia*.' *Episteme*, 6 (1): 52-73.

Sarkeesian, A. 2012. 'TED: The Mirror.' Retrieved March 23, 2013, from http://tedxwomen.org/speakers/anita-sarkeesian-2/.

Simon, J. 2010. 'The Entanglement of Trust and Knowledge on the Web.' *Ethics and Information Technology*, 12 (4): 343-55.

Sober, E. and Wilson, D.S. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Tollefsen, D. P. 2009. '*Wikipedia* and the Epistemology of Testimony.' *Episteme*, 6 (1): 8-24.

Twitter 'FAQs about Verified Accounts.' Retrieved November 19, 2012, from https://support.twitter.com/articles/119135-faqs-about-verified-accounts#.

———. 'The Twitter Rules.' Retrieved February 16, 2013, from http://support.twitter.com/groups/33-report-abuse-or-policy-violations/topics/121-guidelines-best-practices/articles/18311-the-twitter-rules#.

Viégas, F. B., Wattenberg, M., and Dave, K. 2004. 'Studying cooperation and conflict between authors with history flow visualizations.' *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 6 (1): 575-582.

Wallace, K. A. 1999. 'Anonymity.' *Ethics and Information Technology*, 1(1): 21-31.

Whittle, S. 1998. 'The Trans-Cyberian Mail Way.' *Social & Legal Studies,* 7(3): 389-408.

***Wikipedia* contributors**. 'Category: *Wikipedia* Behavioral Guidelines.' *Wikipedia, The Free Encyclopedia.* Retrieved December 20, 2012, from http://en.wikipedia.org/wiki/Category:Wikipedia_behavioral_guidelines.

———. '*Wikipedia*: Counter-Vandalism Unit/Vandalism studies/Study1.' *Wikipedia, The Free Encyclopedia*. Retrieved December 10, 2012, from http://en.wikipedia.org/wiki/Wikipedia:Counter-Vandalism_Unit/Vandalism_studies/Study1.

———. '*Wikipedia:* Protection Policy.' *Wikipedia, The Free Encyclopedia*. Retrieved February 10, 2012, from http://en.wikipedia.org/wiki/Wikipedia:Protection_policy.

Wray, K. B. 2009. 'The Epistemic Cultures of Science and *Wikipedia*: A Comparison.' *Episteme,* 6 (1): 38-51.

York, J. 2011. 'Journalistic Verification, Amina Arraf, and Haystack.' Retrieved January 19, 2013, from http://jilliancyork.com/2011/06/10/journalistic-verification-amina-arraf-and-haystack/.