"I AM IN NO WAY THIS": TROLL HUNTERS AND PRAGMATIC DIGITAL SELF-REFERENCE

MICHAEL LEE HUMPHREY

ABSTRACT

If personae are masks used to communicate a certain character in performance, what happens in rapid unmaskings, especially as they occur in digital space? That question is central to the phenomenon of "troll-hunting". Employing both journalistic and algorithmic tools, troll hunters unmask the offline identity of purveyors of digital hate speech, child pornography, illegal commerce and sometimes just puckish behaviour. Digital citizens have concerned themselves with the efficacy, privacy and ethics of such hunting, but have not as frequently explored another area: the narrative distance between a digital persona and a perceived "real" person behind that persona. Such distances can range from some version of the sentiment, "I am in no way this kind of person" to a comfortable coupling between online-offline selves, even during public shaming. Using textual analysis, I critically examine statements made by those whose digital troll persona were unmasked. I pay special attention to the word 'I' and the dissonance in offline-online personae, long discussed by academics, but also becoming an increasingly practical concern.

KEY WORDS

Internet Persona; Troll-Hunting; Digital Self; Anonymity; Narrative of Self

INTRODUCTION

It appeared, at first, to be cause for celebration, and HanAssholeSolo, a reddit user, basked in it. In early June of 2017, President Donald Trump had posted a video version of an animated GIF on Twitter. Plucking images from a televised event in which pre-president Trump pretended to clothesline a central figure of the World Wrestling Entertainment, the redditor had replaced the face of Trump's original foe with a CNN logo. In a subreddit dedicated to all things Trump, the anonymous member exclaimed, "Wow!! I never expected my meme to be retweeted by the God Emporer [sic] himself!!!" (Romano 2017). But just days later, all of the commentary from HanAssholeSolo, and the GIF itself, were gone. Still anonymous to most of the world, but now known to CNN investigative reporter Andrew Kaczynski, the future of the person behind the GIF rested in the hands of the news organisation he had mocked. In between those events were revelations that HanAssholeSolo was behind many putrid comments about people of colour, women and Jewish people. On the verge of being outed, HanSoloAsshole posted on that same subreddit page:

> First of all, I would like to apologize to the members of the reddit community for getting this site and this sub embroiled in a controversy that should never have happened. I would also like to apologize for the posts made that were racist, bigoted, and anti-semitic. I am in no way this kind of person, I love and accept people of all walks of life and have done so for my entire life. (Kaczynski 2017)

With a single email from a journalist, which made clear that HanAssholeSolo had become the successful target of troll-hunting, the narrative of a digital persona flipped at breakneck speed. While this incident could be studied in many disciplines, troll-hunting's effect on the private-public negotiation of identity in digital space is particularly rich and a useful lens on persona creation for all digital purposes.

Two practical definitions guide this research. *Troll-hunting*, first, is the quintessential Internet vigilante instinct. From journalists to hackers to social media platforms, the goal is simple: find Internet trolls and stop them by blocking access to a platform or exposing (or threatening to expose) the offline person or people responsible. Secondly, what defines *trolls* is negotiated both within online ecosystems (Twitter denizens perceive it differently than redditors) and across social discourse in offline cultures. For a deeper look into the term, Morrissey & Yell's (2016) article presents a useful etymology and history of the term "trolls" as it has emerged in digital space. Edstrom (2016, p. 98) also offers a valuable definition when she writes that "trolls are a metaphor in the Nordic countries that signify beings that fear the light; Internet trolls are people who write offensive things in order to provoke reaction". Phillips and Milner (2017, p. 17) also offer a helpful warning against defining trolling too broadly, using it as a "catch-all" for any mischief or antagonism, which "tends to minimize the negative effects of the worst kinds of online behaviors". So what HanAssholeSolo created that inspired President Trump's tweet might not meet the most rigorous criteria, but many of the statements he made on reddit in other threads did. Yet it was the image that Trump used that made his unmasking big news.

For this paper, I do not aim to plumb the depths of the trolls' gratifications or the effect on digital life. That's not to dismiss the complexities of assuming a separate identity for the purpose of trolling, a field of research that is rich in its own right (Phillips 2011; Milner 2013; Bishop 2013 and 2014; Dynel 2016) or the effects of online discourse, which can be severe (Butler 1997) to the point of life-threatening (Morrissey & Yell 2016). I am interested, however, in the internal relationship between digital troll persona and what is still all too often considered the "real" person behind the digital name and what that tells us about the division between the online and offline self, regularly self-referenced in both spaces as *I*. The recent work of Phillips and Milner (2017), especially their description of "identity play", can speed us forward quickly on this point. Identity play is a fast-moving negotiation between our concept of self and the present or perceived audience, both sides playing critical roles in forming that present-moment identity. Rooted heavily in Goffman's (1959) impression management, the authors argue that these negotiations mean "we all make conscious, unconscious, and sometimes semi-conscious behavioral and linguistic choices to highlight certain masks" (Phillips & Milner 2017, p. 76). We adjust the mask not only to conform to a certain audience, but also sometimes "*toward* a particular audience, *against* a particular object" (p. 79). This is where trolling finds its traction, such as when men target women for the delight of the like-minded and the horror or fury of others, and, "These behaviors, in turn, help create and sustain the *I*; they reveal what a person values, and the groups with which they identify" (p. 78). But this notion of play takes on another set of values, that of the game itself. In the instances I examine below, the light has been shone on the offline people who have stirred the emotions of online communities.

But troll-hunting also snatches away a "win" in the gameplay of identity. If anyone has ever watched professional wrestling in which a masked grappler is involved, you know being unmasked is a greater defeat than being pinned. Though the consequences can be more severe in troll-hunting, the unmasking itself entails a similar sort of victory. This is partially true because creating and sustaining *I* in such a moment is fraught with challenges.

What happens in rapid unmaskings, especially as they occur in digital space? For HanAssholeSolo, a massive rift opens between the pre- and post-hunting *I*, but this is not always the case. In performing textual analysis of digital persona of trolls and their creators, the distance between the pre- and post-hunted troll persona varies widely. Behind that variation is a collection of narratives, which also vary, but illuminate assumptions most users of the Internet appear to hold. In discussing troll-hunting, digital citizens have concerned themselves with the efficacy, privacy and ethics of such searches, but have not as frequently explored another area: the digital dualism (Jurgenson 2011) of considering a digital persona to be constructed by a "real" *I*, which remains prominent among digital citizens.

To resist this common notion, many scholars convincingly address the realness of the online life. I hope to reframe the question by looking in the other direction: be questioning the realness of our offline self, or at least the way we reference that self, via an essay by philosopher G.E.M. Anscombe involving the metaphysics of self-reference. Coupled with the concept of the "dialogical self" by Herbert J.M. Hermans, a unifying theme is described of the offline self as a negotiation of many personae, which I will apply to the responses of trolls when they have been hunted down. I will then conclude this paper by exploring the underlying narratives that allow not only trolls, but most digital citizens, to overemphasise a divide between digital and physical selves.

PERSONA, *I* AND DIALOGICAL SELF

The questions of persona, seen as one side of a delineation of private/public selves, quickly became fodder for early digital theorists. Turkle (1984) famously spoke of a "second self" as an affordance of computing before the World Wide Web existed and Clark (1994, para 5) perceived the persona to be "supplemented, and to some extent even replaced" by the summation of the data available about an individual. He conceives of three types of digital personae—passive (data-creation), active (agentive-presence) and autonomous (digitally self-activating)—which can further be categorised as projected or imposed as well as formal and informal. But his most relevant contribution for this paper is an index of motivations to develop multiple personae in digital space (Clark 1994, para 18):

- the maintenance of a distinction between multiple roles (e.g. psychiatrist or social worker and spouse/parent; employed professional and spokesperson for a professional body; and scout-master and spy);
- the exercise of artistic freedom;
- the experimental stimulation of responses (e.g. the intentional provocation of criminal acts, but also the recent instance of a male impersonating a physically impaired female);
- willing fantasy (as in role-playing in multi-user dungeons and dragons or MUDDs);
- paranoia (i.e. to protect against unidentified and unlikely risks); and
- fraud and other types of criminal behaviour.

Each mask created for digital consumption holds with the core concept of a creator's internal expression negotiating with a public, the "identity play" of Phillips and Milner (2017). So, trolls

might perceive themselves creating a digital persona for "artistic (expressive) freedom" and very likely "stimulation of responses". The data created in that digital performance does not supplement the persona, or replace it, but rather complexifies it. Or, as Nolan (2015) argues, it fragments the self. Using catfishers (pretending to be someone you're not) as a focus, Nolan offers insights into its cousin, the troll, who, "lacks influence and legitimacy because through his similar, but calculatedly covert, interpretation of the fragmented self, he blatantly disregards the social protocols, conscious and subconscious, of lingual identity" (p. 62). To fragment successfully in digital space, the argument continues, the intent must be sincere and the self must be evaluated through a moral lens. Nolan represents a tradition of scholarship that respects the meaningful implications of a digital persona. This tradition rejects the notion of IRL (In Real Life) by asserting the "realness" of online life. As mentioned above, I intend to augment this tradition by taking a slightly different tack.

To do I this, I begin by combining two arguments from Anscombe. In the first, Anscombe (1958) reintroduces, and advocates for, Aristotelian virtue ethics, in which moral decisions are akin to skills developed in interaction with a community that seeks *eudaimonia*, rather than demanding adherence to *a priori* truths. The digital persona of a troll confronts these skills as either absent or prone to subversion, often in contrast to offline persona who might publicly adhere to them. But when the networked public of any, or many, digital platforms resists trolls, the skills needed for a digital *eudaimonia* are naturally highlighted. Offline and online ethical skills, while different in some particulars, align enough that the unmasking of a troll persona can significantly affect the physical world of the bad actor. And here we can see the traditional argument of why online life, and its personae, could be considered "real".

The second, which demands more detail, is Anscombe's concept of first person (1975), in which she addresses the inability to reference anything with the word *I,* a frame to reconsider the "realness" of the offline self or, for that matter, any one self. *I* is, most often, accepted as a meaningful reference in both offline and online expressions of self. For example, readers might accept that *I* is a logical reference in both of the statements from HansAssholeSolo. If both were accepted in that case, we must assume a radical change of heart took place. But Anscombe has a different solution: "'I' is neither a name nor another kind of expression whose logical role is to make a reference, at all" (p. 32). Citing Locke's insight that one who thinks "I did it" might be a "different thinking substance" than the one thinking "I am doing it", she argues that a re-identification of the object must take place each time, thus losing its ability to name. Perhaps, she considers, *I* works like a demonstrative, such as "this" or "that", rather than a name, which succeeds only if it "catches hold", of its object. Descartes claimed it does, but he only meant this in regards to thinking, not a body. Anscombe doesn't buy it, and her answer portends the multiple digital personae we now experience:

> How, even, could one justify the assumption, if it is an assumption, that there is just one thinking which is this thinking of this thought that I am thinking, just one thinker? How do I know that "I" is not ten thinkers thinking in unison? Or perhaps not quite succeeding. That might account for the confusion of thought which I sometimes feel. (p. 31)

So, for Anscombe, *I* is neither name nor any other expression of reference of a person:

> Note that when I use the word "person" here, I use it in the sense in which it occurs in "offences against the person". At this point people will betray how deeply they are infected by dualism, they will say: You are using 'person' in the sense of "body" - and what they mean by "body" is something that is still there when someone is dead. But that is to misunderstand "offences against the

> person". None such can be committed against a corpse. 'The person' is a living human body. (p. 33)

The dualism becomes demonstrable if people suddenly speak of themselves as another person, because it assumes something is not quite right. Anscombe employs William James' imagining of a man named Baldy who falls out of a carriage and bemoans the news that "Baldy" was the one who fell out ("Poor Baldy!") as if he were not the same Baldy. Perhaps he's hit his head and has lost the connection between self and subject, but might still say, "I am sorry for poor Baldy". Anscombe concludes, "[t]he (deeply rooted) grammatical illusion of a subject is what generates all the errors" in thinking about the self (p. 36). And so we find self-identity left on the side of the road, depending on a word, *I,* which has no real reference. All admit, however, that *I* has great function in English, so how can this be?

Rovane's (2004) answer for establishing selfhood connects the dots nicely. She offers two premises for forming, rather than being biologically bestowed with, selfhood: 1) a person is "subject to the normative requirement to achieve overall rational unity within itself"; and 2) a person "must be *committed* to satisfying that normative requirement" (p. 238). This is an agentive notion of personal identity, and a relational one too. She employs research of Dissociative Identity Disorder (DID) as a beginning state from which we attempt to knit together rational unity, a metaphysical given of human nature. She writes:

> Multiplicity is one of the possible states at which human rational activity could be deliberately and coherently directed. When such activity is so directed, it is not, of course, carried from one human-size point of view but from multiple points of view, each of which has separate ends for the sake of which it is striving to achieve rational unity within itself instead of striving for rational unity within the whole human being. (Rovane 2004, p. 248)

This rational unity is a narrative, so argue many personality psychologists (e.g. McAdams 1985; Bruner 1994), that extends over time. If it is true that the use of *I* cannot properly refer to a subject, as Anscombe argues, or consistently refer an object, it could still refer to an action: the *I* is a collection of narrated personae, fragmented, as Nolan suggests. So, *I* is not the subject of "this living thing here", but instead the object of "that unifying act there", that character-being that is described as acting and thinking in the world and acting and thinking text in digital space. There is no fundamental difference in their "reality". Thus, the power of unmasking a troll. That it is not a subject, it seems, is not a problem, because it functions pragmatically in a similar way. *I* identifies an active process of becoming. This does not negate Anscombe's first person argument, but it might render her conclusion less significant. If we agree with the intersection of Rovane and personality psychologists, that humans are generally about the business of unification of the fragmented personae, then it seems both possible and useful to refer to that action as *I.* This can be referred to, not as a body or as a thinker of thoughts, but only an action of a body and its connected brain that engages in a process of unification of components of the narrated self, whether the action is deemed successful or not. Even if the character is not constant, though recognisable, the action is constant. So, in each of the statements—"I am going to the store tomorrow", "I am going to the store right now", and "I went to the store yesterday"—a narration emerges. It is created by the brain, or even algorithms, but is determined (to varying degrees based on the person's ability and desire) by demonstrable things that happen in collectively meaning-making worlds. As Dennett (1992) argues:

> We cannot undo those parts of our pasts that are determinate, but our selves are constantly being made more determinate as we go along in response to the way the world impinges on us. Of course it is also possible for a person to engage in auto-hermeneutics, interpretation of one's self, and in particular to

go back and think about one's past, and one's memories, and to rethink them and rewrite them. (p. 279)

In all of the cases, the *I* is the character in the midst of the action.

But there is still a major problem. If *I* is a process, what is the subject? I would argue it is *I*. Here Kripke (1982) is useful when he refers to Wittgenstein's *form of life*, the set of agreements we create to make words mean something specific. The *form of life* agreement among a community of entities (people or perspectives of one person), would be the subject. It is the community of "selves" that exists in the brain to form unities that we call or accept as *I*. As Rovane (2004) argues, it is not necessary for a full unification to take place for an *I* to emerge— any kind of unity of the fragments in which the community of self agrees to is *I*. For that is the character spoken about in the community. We can understand it simply by thinking about the communal agreement being the subject. If you wanted to get dizzying about it you could say, "[The *form of life* that agrees the community within this body has unified to create a perspective on one element of this body's activity] went to the store yesterday". Or you could just say, "I".

All that is necessary here is agreement of what constitutes the *I* in any given narrative case, but to get agreement you must have more than one, because a beginning state of unity gets us back to the basic problem. To get to a meaningful first-person persona, and not just an empty pronoun, one must have another grammatical person; otherwise there is no community. This is where Nolan and Jurgenson lead us, and where Hermans (1996) can round out the argument. Hermans argues the target is a relational schema (self-to-self, self-to-other and interpersonal script) that drives the "dialogical self", conversations among the many aspects of the self to sustain the narrative of a unified *I*. A dialogical self does not present a "role" for each element of the self but rather a "position", which is voiced to other positions.

The capacity of self-renewal and self-innovation allows the self to engage in an active process of positioning. The use of the verbs positioning and repositioning allows the dialogical self to take initiatives to position itself in new ways, as can be seen in the lives of artists, scientists, and people who renew themselves by breaking at times through the limits of custom and convention. (Hermans 1996, p. 43)

These "breakings" of custom and convention can be conceived positively, but they can also explain how a troll emerges, especially in an anonymous position. When that breaking occurs between the mental unified *I* and digital persona of a troll, the general assumption is that something dark is lurking inside. How the unmasked troll responds to that assumption can be critical to understanding how we, too, have created fragmented personae that start to cause conflicts, either self-to-self or self-to-other, sometimes to tragic effect. To deny that is true for all of us is a dangerous assumption, because our possible negative selves are left unchecked. Perhaps the only mindset more dangerous is to believe that digital personae have little or no power. Troll-hunting offers an unusual opportunity to understand what that disconnecting, or connecting, of physical-digital positions looks like.

NARRATIVES OF UNMASKING

If my initial thoughts about trolling was that the creators' motives were simplistic and cruel, a textual analysis of six recent unmaskings in the U.S. and U.K. at least disabused me of the notion of its simplicity. The narratives of unmasked personae, and how they aligned with the previous troll persona, ranged widely in both expression and content. While I could offer a simple division of *contrite* versus *defiant*, three more interesting categories emerged from the cases.

To perform the analysis, I scraped information from media accounts, social media chatter by and about the troll, and media about the unmasking when it was available. Employing Fairclough (2003), I focused primarily on the multi-functionality of texts, which he categorises in three ways: action (including interpersonal), representation and identification. These three functions work nicely as a frame for persona as each represents a negotiation with publics in digital space. But I also focused heavily on "mediation", which Fairclough cites via Silverstone as the "movement of meaning", transported across "networks" of texts he calls "genre chains":

> ... these are different genres which are regularly linked together, involving systematic transformations from genre to genre. Genre chains contribute to the possibility of actions which transcend differences in space and time, linking together social events in different social practices, different countries, and different times, facilitating the enhanced capacity for 'action at a distance' which has been taken to be a defining feature of contemporary 'globalization', and therefore facilitating the exercise of power. (p. 31)

This is particularly useful in analysing digital media, which now switches from kind to kind effortlessly, from the formal and attainable journalistic approach to informal and, at times, arcane discourse of social media. Journalism often intensifies and reduces discourse due to professional norms, and the journalist's own limited perspective, to create efficient meanings easily consumed. And journalist-induced unmaskings shift the locus of control from the troll to the journalist. Even then, however, the shock of unmasking did not homogenise the narratives among the outed trolls. Each had a particular response to the revelation the he or she had been found out. The responses included three distinct narratives 1) "not really me"; 2) "freedom to speak"; and 3) "trolling as justice". What is interesting in examining each is the relative size of the rift between personae of the unmasked. The "not really me" narrative suggests a large rift between offline and online persona, while the other two present much smaller gaps, but for different reasons.

Not really me

As illustrated above, unmasking of some trolls led to a narrative of quick differentiation between personae, somewhat similar to the classic "devil made me do it" defence. The "devil" in these cases was often the affordance of anonymity. It seems in the statement shared at the beginning of this paper, the man who went by HansAssholeSolo on reddit uses *I* to refer to something approximating a unified self when he says "I am in no way this kind of person, I love and accept people of all walks of life and have done so for my entire life" (Kaczynski 2017). But logic would dictate a rationale for his online actions and this would-be unmasked troll had the impetus to reconsider the effect of toxic digital discourse. He wrote on reddit:

> To people who troll on the Internet for fun, consider your words and actions conveyed in your message and who it might upset or anger. Put yourself in their shoes before you post it. If you have a problem with trolling it is an addiction just like any other addiction someone can have to something and don't be embarrassed to ask for help. Trolling is nothing more than bullying a wide audience. Don't feed your own self-worth based upon inflicting suffering upon others online just because you are behind a keyboard. (Kaczynski 2017)

It is reasonable to suspect that the chance to remain anonymous guided such rapid reflection and repentance.

In a case with far greater consequences, a 24-year-old Isabella Sorley was jailed for twelve weeks for making menacing threats on Twitter (Thomas 2014). After she served the sentence, Sorley agreed to be confronted by one of her victims, a woman named Natalie, in a

meeting arranged by the BBC. This is a vigorous example of Fairclough's genre chains and the rapid shift in power. Because the courts, traditional media and social media interact in this case, the powerless victim of trolling is given the upper hand, with the large media organisation dictating the structure and, ultimately, mediating the engagement for the gratification of an audience. In this situation, the fully unmasked Sorley communicates a similar distance from her troll persona *I* on Twitter. Her narrative also takes on a "devil made me do it" theme, but Sorley uniquely leaves room for uncertainty about the offline-online disconnect. Here is an exchange from the BBC (Thomas 2014):

> Sorley: I'd say I'm quite a nice person.
>
> Natalie: If you're such a good person. Why did you send tweets - including 'Go kill yourself', 'Rape?! I'd do a lot worse things than rape you!!' and 'just got out of prison and would happily do more time to see you buried'?
>
> Sorley: I'd been on an incredibly heavy night out. It was 80p a drink. So, take 20 quid, you're going to be smashed. I can't completely blame alcohol but it's definitely got a part to play in it. I'm a follower, not a leader, and I saw a lot of people were sending those tweets. To say that I'd do worse things than rape is utterly appalling, it's disgusting. I've questioned myself - is there something wrong with my mental state?

Sorley creates clear divisions between three *I's* in this explanation: the offline past self, the online past self and the present self. But, for Sorley, the causes are largely mapped back onto her offline self—drunkenness and a proclivity to follow—rather than using the affordances of digital anonymity as a catalyst. This could be a result of time to reflect, and with that, she evokes a sense of confusion about how two personae connected to her *I* could be so disparate. Here again, we see a capitulation to the closing of a gap between offline and online decision-making with a clear repudiation of one digital persona that Sorley had created on Twitter. These cases, however, proved to be the minority.

Freedom to troll

A very common theme in trolling is either the harmlessness of digital actions and/or the implicit right to post nearly anything as a digital ethic. Not surprisingly then, these are common narratives of an unmasked a troll. One interesting and complex set of reactions, in that regard, comes for a 2012 unmasking by Gawker journalist Adrian Chen, who used deft digital forensics to root out the person behind the reddit troll Violentacrez, known for his "unending fountain of racism, porn, gore, misogyny, incest" (Chen 2012). By the time Chen had built his case to prove that Violentacrez was Michael Brutsch, a 49-year-old programmer for a Texas financial services firm, he was busy informing Brutsch of the impending article to come and seeking comment. The similarities between the Violentacrez story and HanAssholeSolo are overshadowed by key differences: the Gawker journalist was clear on his intent to publicly unmask his troll, while the CNN journalist was ready to keep the offline identity secret based on a few conditions. In that sense, Brutsch is more like Sorley, but without the time to reflect. Still, Brutsch's complete lack of regret is notable, even as he pleaded with Chen to keep his offline identity secret. He argues that he needed to keep his job, because his wife was disabled. He offered to delete his most offensive posts and act as a spy for Chen in reddit's darkest spaces. But when asked if he regretted any of his trolling, the answer was blatant: "I would stand by exactly what I've done" (Chen 2012). The rationale included a mild equivocation on harm ("It's not like I do anything illegal"), but the predominant narrative was freedom.

> He needed to keep his anonymity to protect his ability to express things many people think but hardly anyone says. With Violentacrez, "I got the freedom to talk about my personal life, my personal feelings... I'm sure there's more than one person in this building who's a pervert", he said, referring his office building. (Chen 2012)

That logic reveals a narrative of greater unity between the online and offline *I*, probably because it is largely supported by a large reddit community itself, which forms a twisted kind of ethic. Even before Gawker published the article, more than sixty subreddits had banned the publication. Chen quotes another online platform that covers social media, *The Daily Dot*, to lay out the ethic as it is practiced. Here the real enemy is "doxing", the term of preference of unmasking for the Internet:

> At Web communities like Reddit, which thrive because users are free to say and do anything they want, doxing is a severe crime, both to users and the site's staff. It's far worse than offensive speech like racism and homophobia or, yes, even posting surreptitiously snapped photos of innocent women for creeps to perv over. Why? Because doxing undermines the community's structural integrity: Reddit simply would not exist as we know it if users weren't operating under the freedom of a flexible identity. So redditors aren't banning Gawker to protect violentacrez, they're doing it to protect themselves. (Chen 2012)

Freedom, rather than decency or civility, is the moral that transfers from offline to online. In claiming that, the unified *I* appears to defy convention on both sides of the digital divide.

Trolling as justice

The narratives of trolling as justice and freedom to troll cross over into one another more than once. But the trolling as justice is unique in that it not only defends the troll persona, but also often attacks, again, the target of the trolling to justify itself. That was the narrative of an Internet troll named @sweepyface, who was unmasked by television station Sky News for trolling the McCann family, which had become public figures after their child went missing. In a brief confrontational interview caught on camera, the woman behind @sweepyface, 63-year-old Brenda Leyland of Leicestershire, defended herself by saying she was "entitled" to troll the family (Smith 2014). However, while still fully masked but clearly feeling pursued, @sweepyface narrativised her trolling persona as a force for justice and transparency, bolstered by the belief that the McCanns were involved in their child's disappearance. "I fear that we are in this 4 the long haul, up to all of us to a) Bang home the facts b) make #mccann s live in shame for years" (Smith 2014). When the reporter who would eventually unmask Leyland began to follow @sweepyface on Twitter, she wondered in a tweet why he wouldn't "investigate some of these facts and show neutrality".

This seemed, for a time, to fit neatly in the category of trolling as justice, but a brutal twist belied Leyland's narrative. Just two days after Sky News aired the report about her, Leyland was found dead in a hotel room in a Leicester hotel room, a death later determined to be suicide (Davies & Conlan 2015). In the inquest, it was clear that @sweepyface's persona was no match for Leyland's offline persona in the battle for a unified *I*. In a conversation with the reporter before the story aired, she told him, "Oh, I've thought about ending it all but I am feeling better. I have had a drink and I've spoken to my son". While the inquest made it clear that Leyland suffered from depression, and had attempted suicide before, her son wrote in a statement that was read at the inquest that Leyland "could not bear to think she could be disliked by those in her community" (Davies & Conland 2015).

In a very different outcome, also in England, Twitter troll @Holbornlolz spun a narrative of defender of liberty and attacker of institutions, as well as a digital satirist. What distinguishes this narrative is the timeframe—he made one part of the argument during the unmasking confrontation, when it became clear he was 51-year-old Robert Ambridge of Essex, and another part of the argument two years after being unmasked (Murfitt & Luck 2013; Daubney 2015). Here the power differential is more unclear: who is getting the best of whom in the exchange? On the one hand, the reporters are getting behind the digital persona, and arguably exposing the offline persona's justifications. On the other hand, Ambridge could arguably be normalising the act of trolling in the broader culture. In the moment of realisation that he had been unmasked, Ambridge focused most of the harmlessness ("entertainment") of his actions, which included making fun of people's bodies and using tragedies such as the Boston Marathon bombing as fodder for his "satire". He told the reporters who unmasked him: "This is dark humour. People might not like my humour but I think it is funny and it gets a chuckle" (Murfitt & Luck 2013). In that statement, the offline *I* seems to pull the online *I* closer, while simultaneously objectifying it with praise and acknowledgement of enjoyment from others. Two years later, even after Ambridge was forced to move to another town, his narrative had become more sophisticated and more tied to justice: "I'm here to expose the hypocrisy of it all. I despise politicians, their endless lies, their assumed authority and the constant interference of 'the State' in how I choose to live my life" (Daubney 2015). But it wasn't only offline institutions he intended to check, it was also the conventions of the Internet itself.

> Now let's be honest here, most people adore being outraged. It gives them the aura of moral superiority and they can parade their smugness for all to see and judge them by. They can wear it as a badge to indicate their adherence to 'better standards' as if that gives them the right to silence anyone who doesn't measure up to their ideals. (Daubney 2015)

What is clear in the juxtaposition of these two cases is that both online and offline personae can spin similar narratives while the distance between a perceived *I* and the digital persona varies greatly. It is also clear that distance does not necessarily determine the actions of the digital persona—both can spout hateful speech, but the unmasking of the digital troll can have very different consequences. Morrissey and Yell (2016, p. 29) examine this briefly in light of the Leyland case writing, "[t]he relation between the public persona of the troll as manifested in their online discourse, and their private selves is complex and apparently contradictory". In my analysis, I found that "apparently contradictory" depends on the troll him or herself.

BEYOND AUTHENTICITY AND TOWARD COMMUNITY

The premise of this paper was to observe the responses about the distance between digital and offline *I*'s when trolls are unmasked by troll-hunters; not as a form of schadenfreude or even to better understand trolls, but as a lens into all personae in digital space. Wrested from the subject of trolling, the three themes of why we divide our online-offline personae still hold: the online self is "not really me"; 2) we have the freedom to act differently online; and 3) what we do online is just. Not surprisingly, the variation was as great as the masks themselves, but there are possible extrapolations. In this same journal, McRae (2017, p. 25) makes a statement that resonates with the identity play of Phillips and Milner (2017): "To some extent, we all craft personas with a real or imagined critical audience in mind". How we wish to affect those imagined audiences is a meaningful difference between most of us and trolls. However, the crafting of the persona as something different than our offline personas likely has fewer differences. It is still common, and understandable, to accept some "freedom" in crafting a digital persona. Digital life affords "shaping" of a self, "no one knows you're a dog", as the

cartoon trope goes, but there's a point missing in this. That point refers all the way back to a realisation that the collection of digital personae that populate social platforms come with different expectations and predispositions. In seeking or encountering audiences for the "game" of identity, users soon find themselves negotiating (or flat-out fighting) for virtues of "how to be" in digital space. That some of these spaces feel encapsulated from a larger society, digital or physical, would only reinforce the sense of freedom. In this negotiation, Phillips and Milner (2017, p. 85) convincingly argue the role of anonymity is not necessarily predictive of one form of behaviour: "In short, deindividuated, anonymous participation online can facilitate the bad, the good, and the in-between, resulting in every permutation of communicative expression imaginable". This freedom, like all real freedom, takes on different forms. But what the unmasking of trolls demonstrates is how quickly the communicative expression can change when implications from two different aspects of the self collide.

This is clear even in the most rigorous defence narratives of trolls. Shame is just as powerful, if not more so, than confirmation and sycophancy. The consequences for "breakings" from a generalised norm can feel as meaningful online as it does offline, and that's assuming that the consequences don't cross-contaminate the other. McRae (2017, p. 14) deftly explains why, despite these consequences, some trolls I examined (like any other social media persona) hold closely to their narrative of freedom or justice to resist judgment from others:

> When publics perceive evidence of unoriginality and inconsistency in [...] personas, they are less likely to accept personas as authentic, recognizing instead that personas are constructed.

Had Violentacrez responded the way HanAssholeSolo did, he would have likely met the same digital fate: scorn and obliteration from his own community. It seems clear from Violentacrez's narrative that remaining "authentic" to his community of trolls was more important than becoming palatable to his offline community, such as his workplace, which fired him. The idea that one set of personae (the offline set) would have many important life implications, while the other set (online) would live free of those implications is a naiveté that has been exposed for years. In referring to cybersex, for example, Dibbell (1993) made this point a quarter of a century ago:

> To participate, therefore, in this disembodied enactment of life's most body-centered activity is to risk the realization that when it comes to sex, perhaps the body in question is not the physical one at all, but its psychic double, the bodylike self-representation we carry around in our heads. I know, I know, you've read Foucault and your mind is not quite blown by the notion that sex is never so much an exchange of fluids as it is an exchange of signs. But trust your friend Dr. Bombay, it's one thing to grasp the notion intellectually and quite another to feel it coursing through your veins amid the virtual steam of hot netnookie.

Jurgenson (2011) extrapolates this one experience to all digital experiences when he contests the concept of a "second self" as digital dualism and argues that all of the experiences live in one sphere rather than two: "[w]e are not crossing in and out of separate digital and physical realities, ala The Matrix, but instead live in one reality, one that is augmented by atoms and bits". Mine is less an argument to yet again defend the "realness" of online life and more an honest look at the construction of self through all aspects of a life. "In short, we all deceive, on the internet and in our own living rooms" (Phillips & Milner 2017, p. 81). Even beyond deception of others, we daily deceive ourselves in constructing an *a priori* version of *I.* Postmodernist thought, and much of personality psychology, has also travelled down this road for decades now. On a practical level, however, we continue to see daily signs that the

commonality among an individual's personae, across all related *I*'s, is lost on many when they log in to digital platforms. The unifying theme of digital self, anonymous or not, is less a certain way to be, but the dualism itself.

Werning (2017) turns the common response, prosecuting the mask as inauthentic, into prosecuting the notion of reality as it is often perceived. In that light, the mask is what is authentic. In a postmodern world, this seems particularly wise. On the other hand, I am sceptical of McRae's (2017) solution of authenticity as labour, at least as it applies to negotiating the moral landscape of public digital life *vis-a-vis* trolls. Unmaskings have taught us that the range of authenticity varies widely among trolls. Instead, I would like to return to Anscombe's virtue ethics as a possible lever from which to activate a digital *eudaimonia*. It is sometimes lost that digital persona is a very new phenomenon, especially compared to ancient concepts of persona. Just because persona acts similarly in digital space, and often evokes similar consequences, does not mean human beings have gained proficiency in learning to act morally as digital personae, yet. Virtue ethics argues that building an atmosphere for thriving means developing social agreements and honing individual virtue. There is no clear right or wrong in digital space unless desired outcomes are widely agreed upon. To be authentic seems too individually constructed, both in creating and judging its presence. To create a space where the most people thrive reflects an age-old practice of building *eudaimonic* cultures. That's not to discount authenticity as whole, but rather shifts the focus on developing strong, positive relationships for the good of the whole digital ecosystem. "I am in no way this" is less a division and more a statement of becoming within a system that would foster personae that benefit communities. If this sounds naive, so it is, if we think that a majority of personae wanting such a digital *eudaimonia* naturally produces it. And that might offer one more lesson from all three categories of trolls and the troll-hunters who unmasked them. While there is a range of ethical modifiers that might be used for trolls, from vile to problematic to mischievous, there is a neutral attribute they share that is generally worth stealing: most are vocal about creating the digital world in which they want to be.

WORKS CITED

Anscombe, GEM 1958, 'Modern moral philosophy', *Philosophy*, vol. 33, no. 124, pp. 1–19.
— 1975, 'The first person', in SD Guttenplan (ed.), *Mind and language*, Oxford University Press, Oxford, pp. 45–65.
Bishop, J 2013, *Examining the concepts, issues, and implications of Internet trolling*, IGI Global, Hershey.
— 2014, Digital teens and the 'antisocial network': Prevalence of troublesome online youth groups and internet trolling in Great Britain. *International Journal of E-Politics,* vol. 5, no.2, pp. 1–15.
Chen, A 2012, 'Unmasking reddit's Violentacrez, the biggest troll on the Web', *Gawker*, 12 Oct, retrieved 2 Aug 2017, http://gawker.com/5950981/unmasking-reddits-violentacrez-the-biggest-troll-on-the-web.
Clarke, R 1994, 'The digital persona and its application to data surveillance', *The Information Society*, vol. 10, no. 2, pp. 77–92.
Daubney, M 2015, 'Britain's vilest troll: "I'm here to expose hypocrisy', *The Daily Telegraph*, 5 Feb, retrieved 2 August 2017, http://www.telegraph.co.uk/men/thinking-man/11390745/Britains-vilest-troll-Im-here-to-expose-hypocrisy.html.
Davies, C & Conlan, T 2015, 'Woman killed herself after being doorstepped over McCann trolling', *The Guardian*, 20 Mar, retrieved 7 Sept 2017, http://www.theguardian.com/media/2015/mar/20/sky-news-mccann-brenda-leyland.
Dennett, DC 1992, 'The self as a center of narrative gravity', in F Kessel, P Cole & D Johnson (eds), *Self and consciousness: multiple perspectives,* Erlbaum, Hillsdale, pp. 275–288.

Dibbell, J 1993, 'A rape in cyberspace: how an evil clown, a Haitian trickster spirit, two wizards, and a cast of dozens turned a database into a society', *The Village Voice*, 23 Dec, retrieved 19 Oct 2017, http://www.juliandibbell.com/texts/bungle_vv.html.

Dynel, M 2016, '"Trolling is not stupid": Internet trolling as the art of deception serving entertainment', *Intercultural Pragmatics*, vol. 13, no. 3, p. 49.

Edstrom, M 2016, 'The trolls disappear in the light: Swedish experiences of mediated sexualised hate speech in the aftermath of Behring Breivik', *International Journal for Crime, Justice and Social Democracy*, vol. 5, no. 2, pp. 96–106.

Fairclough, N 2003, *Analysing discourse: textual analysis for social research*, Routledge, New York.

Hermans, HJM 1996, 'Voicing the self: from information processing to dialogical interchange', *Psychological Bulletin*, vol. 119, no. 1, pp. 31-50.

Jurgenson, N 2011, 'Digital dualism versus augmented reality', *Cyborgology,* 24 Feb, retrieved 19 Oct 2017, https://thesocietypages.org/cyborgology/2011/02/24/digital-dualism-versus-augmented-reality/.

Kaczynski, A 2017, 'How CNN found the reddit user behind the Trump wrestling GIF', *CNN*, Available at: 5 Jul, retrieved 2 Aug 2017, http://www.cnn.com/2017/07/04/politics/kfile-reddit-user-trump-tweet/index.html.

Kripke, SA 1982, *Wittgenstein on rules and private language: An elementary exposition*, Harvard University Press, Cambridge.

McAdams, DP 1985, *Power, intimacy, and the life story*, The Guilford Press, New York.

McRae, S 2017, '"Get off my Internets": How anti-fans deconstruct lifestyle bloggers' authenticity work', *Persona Studies*, vol. 3, no. 1, pp. 13–27.

Milner, RM 2013, 'Hacking the social: Internet memes, identity antagonism, and the logic of Lulz', *The Fibreculture Journal*, no. 22, retrieved 2 Aug 2017, http://twentytwo.fibreculturejournal.org/fcj-156-hacking-the-social-internet-memes-identity-antagonism-and-the-logic-of-lulz.

Morrissey, BC & Yell, S 2016, 'Performative trolling: Szubanski, Gillard, Dawson and the nature of the utterance', *Persona Studies*, vol. 2, no. 1, pp. 26–40.

Murfitt, N & Luck, A 2013, 'Father of six who is one of Britain's vilest internet trolls exposed after posting sick jokes about Hillsborough disaster', *Mail Online*, 7 Sept, retrieved 7 September 2017, http://www.dailymail.co.uk/news/article-2415122/Twitter-troll-unmasked-Father-Britains-vilest-internet-trolls-exposed-posting-sick-jokes-Hillsborough-disaster.html.

Nolan, MP 2015, 'Learning to circumvent the limitations of the written-self: the rhetorical benefits of poetic fragmentation and Internet "catfishing"', *Persona Studies*, vol. 1, no. 1, pp. 53-64.

Phillips, W 2011, 'LOLing at tragedy: Facebook trolls, memorial pages and resistance to grief online', *First Monday,* 5 Dec 2011, retrieved 19 Oct 2017, http://firstmonday.org/ojs/index.php/fm/article/view/3168.

Phillips, W & Milner, RM 2017, *The ambivalent Internet: mischief, oddity, and antagonism online*, John Wiley & Sons, New York.

Romano, A 2017, 'Trump's anti-CNN tweet originated from Reddit's largest right-wing extremist forum', *Vox*, 3 Jul, retrieved 1 Sept 2017, https://www.vox.com/culture/2017/7/3/15913966/trump-cnn-wrestling-tweet-the-donald-origins.

Rovane, C 2004, 'A nonnaturalist account of personal identity', in M. De Caro & D. Macarthur (eds.), *Naturalism in question*, Harvard University Press, Cambridge, pp. 231–258.

Smith, P 2014, 'The deleted tweets Brenda Leyland sent about the McCanns before she died', *BuzzFeed*, 6 Oct, retrieved 2 Aug 2017, https://www.buzzfeed.com/patricksmith/read-the-deleted-tweets-brenda-leyland-sent-the-mccanns.

Thomas, E 2014, 'Twitter troll: what I said was utterly appalling and disgusting', *BBC*, 16 Nov, retrieved 2 Aug 2017, http://www.bbc.co.uk/newsbeat/article/30075370/twitter-troll-what-i-said-was-utterly-appalling-and-disgusting.

Turkle, S 1984 *Second self*, Simon & Schuster, New York.

Werning, S 2017, 'The persona in autobiographical game-making as a playful performance of the self', *Persona Studies*, vol. 3, no. 1, pp. 28–42.