

## Speech, Mockery, and Sincere Concern: An Account of Trolling

Jeffrey Kaplan

This paper offers an account of a phenomenon that seems increasingly common in public discourse: trolling. The term “troll” is colloquial, and no formal synonym exists in English. But the informality of the term should not mislead us into thinking that the underlying concept is so unimportant as to be unworthy of philosophical attention or so ill-behaved as to be resistant to philosophical analysis. This paper presents such an analysis.

In 2007, several news outlets, including Fox News Channel, began reporting on a new drug that was becoming alarmingly popular with American teenagers: Jenkem. It was made from several ingredients, including human waste, and it produced a powerful high when inhaled. News reports urged parents to look for tell-tale signs of Jenkem-use and to warn their children about the drug.<sup>1</sup>

Jenkem, however, never existed. It was a fictional drug, designed in a coordinated effort by users of the now-defunct website, TOTSE, with the specific intention of provoking the overwrought concern of the American news media.

The Jenkem ruse was, so to speak, a ‘prank.’ But it was not the kind of simple prank, like ringing someone’s doorbell and running away, that seems to principally involve amusement at someone else’s misfortune. The Jenkem ruse—as we will soon discuss in greater detail—seems to have been focused not so much on the *misfortune* of traditional news media and concerned parents, but on their *reaction*.<sup>2</sup> Although it is hard to say precisely what the perpetrators of the Jenkem ruse had in mind,

---

<sup>1</sup> Phillips 2015b.

<sup>2</sup> Though I think it is appropriate to call this particular instance of trolling a ‘prank’, it is hard to say whether *all* trolling is a form of pranking. Part of the reason for this difficulty, of course, comes from the fact that pranking is itself somewhat mysterious. There is, however, some etymological evidence for a connection. Some of the earliest English use of the word “trolling” comes from the early 1970s when US Navy pilots would go “trolling for MiGs”, which was the use of “decoys, with the mission of drawing” enemy fire away from sensitive US military outposts (Saar 1972). In contemporary Japanese, *tsuri* (釣り) and in Korean, *nak-si* (낚시) both literally mean *fishing* and are used very similarly to the English word “trolling.” In Thai, *krían* (เกรียน) literally refers to a hairstyle associated with school-age boys in Thailand is used to refer to behavior similar to the behavior that would be characterized in English with the word “trolling.”

we can imagine a case in which they intended to evoke alarm, concern, or panic. In that scenario, we have a word for this type of behavior in 21st-century English: “trolling.”

The aim of this paper is to offer a philosophical account of this widespread feature of public discourse and to draw some limited practical conclusions from that account. The term “trolling” is the only label that currently exists in the English language for this feature of public discourse, and so it is the one that I use. It is a colloquial term. But the lack of a more formal designation should not mislead us into thinking that the underlying phenomenon is unimportant or unworthy of philosophical attention. In the following, I present a set of conditions that are jointly sufficient and individually very close to necessary.<sup>3</sup> However, in an effort to favor an account that is illuminating over one that is extensionally perfect, there is also a brief discussion of some characteristics of trolling that are not essential, but which nonetheless seem paradigmatic. Once the account has been presented, it becomes clear, for instance, that though trolling is made more prominent by certain features of modern life, it is not new. It is as old as almost any object of philosophical interest.

Before presenting the account of trolling, it is worth briefly addressing two concerns that some readers may have from the outset. First, since this paper does not offer a moral assessment of trolling, is it impossible to provide an account? The idea here is that *trolling* might appear, to some, to be a morally-loaded, thick normative concept. As such, it would be impossible to independently identify trolling, or even individual instances of trolling, without making moral judgments. Luckily, however, *trolling* is not a thick normative concept. We can see this by noting that act consequentialists are unable to make a moral assessment of an instance of trolling without knowing its consequences, but they seem able to classify such an instance as trolling without that information. If this is right, and *trolling* is not a thick normative concept, then trolling may still be morally bad. It may even be necessarily morally bad. But if this is so, it is not as a matter of *conceptual* necessity.<sup>4</sup>

The second concern can be expressed as follows: isn’t the English use of “trolling” so loose and fluid so as to make an account of the underlying concept impossible? This concern is especially pressing if we take “trolling” to be a vague, broad label for internet bullying.<sup>5</sup> That, however, is not

---

<sup>3</sup> It might be thought that *trolling* is a family resemblance or prototype concept, not one structured in the form of necessary and sufficient conditions and that therefore this approach is misguided. Section 4 focuses on non-necessary but still paradigmatic features of trolling, and in that way the concept is treated as a family resemblance concept. But also, as has been noted before, family resemblance concepts can only be analyzed by specifying the specific respects in which instances are similar to the prototype—since not all respects of resemblance, such as time, location, etc., are as relevant as others—and this specification is not so different from providing necessary and sufficient conditions.

<sup>4</sup> Coons & Weber (2014, 6) make a similar point in regards to a similar potential objection leveled against the possibility of a descriptive account of the concept of *manipulation*.

<sup>5</sup> This is the first appearance in this paper of the word “internet.” My understanding is that in standard academic practice is to capitalize the term in its noun form (e.g. “...on the Internet.”), but not in any adjectival form (e.g., “an internet troll” or “part of internet vernacular”). I follow this convention here.

how I understand the term here. The target of this paper is a more circumscribed concept.<sup>6</sup> As to the more general worry that even the more restricted notion of *trolling* cannot be helpfully captured with a philosophical account, the proof of the pudding will be in the eating. If the account developed here meets with enough of our judgments about what is and is not trolling, and if it seems to strike at the heart of the matter, then that is itself strong evidence that there is a single, well-behaved concept that we share and that many of us express with the noun and verb forms of “troll.”<sup>7</sup>

## 1 Starting Points

Here are four plausible pre-theoretic claims about trolling. Any one of them may turn out, upon further inspection, to be false. But it is nonetheless helpful to begin with some partially fixed points against which to judge proposed accounts of trolling.

(A) Trolling can consist of either a single speech act or a sustained communicative interaction. This pre-theoretic claim, as I have stated it here, presupposes that trolling consists of speech or other communicative acts. This is an artificial simplification. It may turn out that wholly non-linguistic behavior can constitute trolling.<sup>8</sup> However, once we restrict ourselves to linguistic cases, it appears clear that trolling can consist either of single or multiple communicative acts. As it turns out, the Jenkem ruse involved some sustained interaction whereby the trolls responded to news outlets with additional ‘data’ and photos of the supposed drug. But, plausibly, the ruse would still have been trolling had it consisted of, for example, a single internet post. In light of this, it will count in favor of any putative account of trolling if that account allows both singular and sustained speech acts to be instances of trolling. However, this consideration, like all the others that follow, is not dispositive.

(B) Trolling is fun (for the trolls).

---

<sup>6</sup> Of course, a considerable amount of trolling occurs on the Internet and much of it constitutes bullying. But the target of this paper is neither the most general nor the most vicious concept expressed by the English word “trolling.” Rather, it is a concept that, unlike *bullying* or *harassment* or *hostility*, has a central relation to mockery. Moreover, bullying often aims to silence and intimidate, which, as we will see, is nearly *antithetical* to trolling as it is understood here. For those readers who inextricably associate the word “trolling” with internet bullying, it may be helpful to think of the word “trolling,” as used here, as a generic label—as if we had designated the phenomenon with “X” or some other variable—and instead focus on the examples used throughout the paper to get a tentative grasp on a phenomenon that will more slowly come into focus.

<sup>7</sup> It is worth briefly mentioning something about the noun. The term “troll” is used here to refer to anyone who engages in even a single instance of trolling, and not merely to habitual or serial trolls.

<sup>8</sup> One example of non-linguistic trolling may be a hedge maze carved by a gardener in such a way that those walking through the maze could never reach a finishing point, always being led back to the start. The gardener carved this maze in order to laugh at the frustrated complaints of those walking the maze.

Trolling is a form of voluntary behavior. And voluntary behavior can be motivated by things other than enjoyment: a feeling of moral compulsion, interest in some likely future reward, etc. So it could turn out that trolling is not fun for the trolls. But trolls do *seem* to find trolling fun. So it counts in favor of a putative account if it allows that trolling might be fun for the trolls, and it counts even more so in favor of an account if it partially explains *why* trolling is fun.

(C) Trolling is frustrating (for those being trolled).

Of course, it can be much more than frustrating, often constituting a serious harm. But, as several examples we consider below demonstrate, trolling can be less damaging. I suspect that almost all potential accounts will have an easy time accommodating this pre-theoretic claim about trolling. Still it counts in favor of such an account if it allows that trolling is often cruel and harmful, sometimes harmless, but nearly always frustrating.

(D) Trolling is made easier and more popular by social media and the Internet.

This could also turn out to be false. But barring any surprising reasons to think otherwise, an account that appropriately connects certain features of social media and the Internet to trolling has at least one point in its favor.

## 2 An Initial Account of Trolling

Consider the following two-condition account of trolling. To troll is to perform an utterance or series of utterances that:

- (1) is likely to elicit alarm or outrage, and
- (2) is believed by the speaker to be likely to elicit alarm or outrage.<sup>9</sup>

This yields the incorrect result that those who knowingly elicit outrage as a mere byproduct—such as biblical prophets who proclaim that the people are immoral sinners—are trolling. So we should add that a trolling utterance or series of utterances:

- (3) is primarily intended to elicit alarm or outrage.<sup>10</sup>

---

<sup>9</sup> The disjunction in each of these conditions may seem odd. Is there something that unifies these two similar, but distinct reactions and explains why they are grouped together in this account of trolling? Are there other reactions that could be plausibly added to this account of trolling? These questions are answered very directly in section 2.3.

<sup>10</sup> Notice what condition (3) is *not*. It is not a requirement that the utterance is insincere. That would have been the most straightforward way to improve the theory if, after all, what made the biblical prophet not a troll was that she believed what she said. But this is too strong a requirement because one can troll by making a statement that one believes to be true.

This intention need not be fulfilled—i.e., the intended outcome need not occur—in order for the utterance to constitute trolling.<sup>11</sup> But the outcome cannot merely be a predictable byproduct.<sup>12</sup>

This three-part account will need to be revised. But briefly considering the account before making the necessary modifications ends up being a more perspicuous way of presenting the advantages of accounts of this broad type. To start, according to the account just presented, the Jenkem ruse is trolling. Or rather, since conditions (2) and (3) concern the inner workings of the speaker’s mind, the account yields a plausible conjecture to that effect. This uncertainty, however, is a *virtue* of the account. Notice that the account presented so far is not uncertain in its application to the Jenkem case if we *stipulate* that the perpetrators of the ruse have the belief and intention specified in conditions (2) and (3). But when we are discussing the real-world Jenkem case, we can only make educated guesses as to the speaker’s mental state. So our account only yields an educated guess as to whether the Jenkem ruse is trolling. Ordinary English speakers are unsure whether someone, or some group of people, is trolling—as contemporary life often involves asking oneself, so to speak, “are they for real or just trolling?”—in precisely those cases where they lack the information about the speaker’s knowledge and intentions that makes the application of the account uncertain.<sup>13</sup>

## 2.1 Advantages of this Account

The first advantage of this account is that it suggests an explanation for pre-theoretic claim (B), that trolling is fun for the trolls. When one causes others to react a certain way, one exerts power over them.<sup>14</sup> I do not have a worked-out philosophical account of the kind of ‘power’ at play, but I do not think I need such an account. By ‘power’ I mean only a kind of purposeful influence over the behavior of others. The suggestion that this kind of power can be pleasant for those who wield it is, I think, rather plausible. Consider the following, from the Scottish comedian, Brian Limond:

[O]ne night on Xbox Live, a group of English gamers took the piss out of my Scottish accent, then buggered off laughing. A new lot

---

<sup>11</sup> Trolling is not a perlocutionary act and, correspondingly, “troll” is not a success verb. A similar point has been made about many verbs including “lie” (see Mannison 1969, 135; MacCormick 1983, 9). One can “unsuccessfully troll,” but this means merely that one trolled but failed to generate the intended reaction.

<sup>12</sup> This way of thinking has some affinity to a well-known passage concerning “disputes with men” from Hume’s second enquiry (2006, sec. I). According to the account settled on in the following section, the behavior that Hume describes here is not trolling.

<sup>13</sup> So, to make this point clear if it is not already, the account presented so far can be applied with determinacy when all of the relevant information is provided or stipulated. It’s application is only uncertain when that information is withheld. If we assume that the concept *trolling* being used by ordinary speakers is the one revealed by this sort of conceptual analysis, then the epistemic position of the philosopher is the same as the epistemic position of the ordinary speaker. That is, given a fixed set of information about a case, we would be precisely as uncertain as to whether the concept applies as an ordinary speaker is. And, as a virtue of this account, we are.

<sup>14</sup> In connection to this, Sartre (1948) characterises the anti-Semite as exhibiting behavior that we would recognize as trolling and as being motivated by a desire by those in the middle class to claim power over their nation.

came along and did the same. Anything I replied with was repeated back to me in the voice of Scotty from Star Trek. I couldn't win. I felt so powerless, so weak, so Scottish. Until I started trolling. When the next lot of English came along, I put on a laidback American accent and said, "Speak English, you dumbass Brits, speak the language we taught you." As you can imagine, they went mental: "It's not your language, you obese, American tit, why d'you think it's called English?" My reply? "I can't understand a word you are saying, you dumbass Brit. English, please. Speak English..." I kept it going for hours, different games, different players, sometimes adding in the odd reference to the Queen and saving their ass in the war, right up until 5am the next day. And I can tell you, hand on heart, that it was one of the best nights of my life.<sup>15</sup>

This is about as straightforward of an example of trolling being used by the disempowered to feel a pleasant power over others as one is likely to find. My suggestion is that this sort of motivation is ubiquitous, even in cases where such power dynamics are less obvious or explicit.

If trolling is pleasant in this way, then it is attractive to those who *feel* powerless, who are only sometimes those who *are* powerless.<sup>16</sup>

The account also suggests that trolling is not as new as it sometimes appears. One suspects, for instance, that teenagers have been trolling their parents for quite some time. And the account explains why trolling would thrive within the child-parent relationship. A teenager is just the right age to feel entitled to power, but be annoyingly kept under the authority of her parent—increasing the benefit of trolling. And, in a sufficiently loving home, a teenager knows that she will still be loved even after saying outrageous things—decreasing the cost of trolling. Similarly, the account explains pre-theoretic claim about trolling (D)—that the Internet is particularly conducive to it. There is a great deal of empirical evidence that the anonymity and geographical distance made possible by the

---

<sup>15</sup> Limond 2012.

<sup>16</sup> As the expression goes: when one is accustomed to privilege, equality feels like oppression. There is some empirical evidence that men are more likely than women to troll, or at least to engage in anti-social online behavior (see Ferenczi, Marshall, and Bejanyan 2017). This fits with the claim that those who feel disempowered are more attracted to trolling because even if men are, on average, more socially powerful than women, their relative power has diminished in the late 20th and early 21st centuries, which is plausibly associated with a feeling of powerlessness.

Internet leads to abusive or outrageous behavior.<sup>17</sup> Plausibly, this is achieved by reducing the social costs of such behavior, while leaving the pleasant feeling of power over others undiminished.<sup>18</sup>

So, on this account, though the Internet is particularly conducive to it, trolling exists independently of, and prior to, the Internet. This point is made by Rachel Barney in her 2016 paper, [*Aristotle*], *On Trolling*, which is a conceptual analysis of trolling written in the style of Aristotle:

[I]t is clear from this that there can be trolling outside the internet. For every community of speakers holds certain goods in common, and with them the conversation [*dialegesthai*] as an end in itself; and the troll is one who seeks to damage it from within. So a questioner can troll a political meeting, and academics troll each other in committees when they are bored; and a newspaper columnist may be a profit-troll towards a whole city.<sup>19</sup>

---

<sup>17</sup> Suler (2004) shows that social inhibitions are significantly reduced when some form of interaction is conducted online as opposed to in person, though it could not be determined if this was because of the anonymity of online interaction or because of other features of online interaction. Suler does note that anonymity allows for cruel behavior “almost as if superego restrictions and moral cognitive processes have been temporarily suspended from the online psyche.” There is some evidence that the disinhibition of online communication comes partly from the lack of nonverbal behavioral cues pervasive in face-to-face interaction (Boudourides 1995) and also from anonymity itself (Hardaker 2010). Douglas and McGarty (2001) did investigate the role of anonymity and found that those whose online identities were anonymous were much more likely to engage in abusive behavior than those whose identities were non-anonymous. Kiesler et al. (1984) also found the same relationship between anonymity and derogatory comments. Then, of course, there is a more general psychological literature concerning not internet comments in specific, but the general effects of “deindividuation” on behavior. The classic deindividuation experiments showed that children trick-or-treating on Halloween in large groups and with masks were nearly ten times more likely to steal extra candy or money than if children lacked masks and were not in large groups. The same effect is present when one feels free to scream at another motorist from within one’s own car.

<sup>18</sup> In addition to the evidence that anonymity increases antisocial behavior, there is empirical evidence that those who feel frustrated or put down are more likely to troll. A recent study (Cheng et al. 2017) attempted to measure how mood and self-esteem affect anonymous commenting behavior. Subjects who were given a difficult quiz, along with results claiming that their performance was below average, were considerably more likely to write cruel or derogatory comments on an unrelated news article than those given an easy quiz and results claiming that they had done well. The study involved both the subjects’ self-esteem and anonymity, but only the former was the investigative target of the study. That is, the difference between the experimental and the control group was the difficulty of the quiz and the subsequent results presented to the subjects. So this randomized, controlled trial was meant to measure the effect of self-esteem on commenting behavior. Anonymity, on the other hand, was simply a feature of the commenting conditions in both groups, so this study itself does not suggest anything about how anonymity does or does not affect commenting behavior (though there is other empirical evidence suggesting that the effect is positive and significant, *supra* note 18). Also, of course, this study measures cruel internet behavior, not trolling.

<sup>19</sup> Barney 2016, 194.

As just one specific example of trolling entirely outside the Internet, in 1978, Mark Gubin painted the message “Welcome to Cleveland” in enormous letters on the roof of his art studio, which was visible to airplanes approaching the Milwaukee airport.<sup>20</sup>

## 2.2 A Less Vicious Troll

Despite its many advantages, this account yields the wrong result in regard to someone who is often considered the greatest internet troll alive: KenM.

Kenneth McCarthy is a 30-something copywriter living in Brooklyn, New York. He comments on various internet sites under the now-famous username, “KenM”. Below is an exchange with the sausage-producing subsidiary of Tyson Foods, Jimmy Dean Sausage. The entire exchange takes place on the Jimmy Dean Facebook page, which is publicly visible to anyone with a Facebook account [*sic* throughout].

*KenM:* i was suppose to get a free sample of your sausage product but they ran out

*Jimmy Dean Sausage:* Hi Ken - Thanks for getting in touch. Can you please send us a direct message (located under our cover photo) with your contact information so we can help?

*KenM:* thank you please send the sausage to 3250 Beulah Rd Pensacola FL 32526

*Jimmy Dean Sausage:* Hi Ken - Can you send us a direct message? The link to send us a direct message is located under our cover photo?

*KenM:* {please put this as a direct msg} well they were handing out pieces of your sausage at walmart but they ran out when I was next in line and some folks took more than 1 piece

*Jimmy Dean Sausage:* Hi Ken - Can you send through your phone number, please? Thanks so much!

*Ken M:* Thank You i sent my number but have not received a call

*Jimmy Dean Sausage:* Hi Ken - You only provided us with your address. Can you share your number as well? Thanks so much!

*KenM:* Thank You our phone # is a landline so we had to send it to your headquarters in the regular old mail.

---

<sup>20</sup> An additional point about this trolling and the Internet: the structure of some internet communication platforms are more conducive to trolling than others, and the account being developed can explain this fact. Twitter, for instance, is rife with trolls. Aside from anonymity, which reduces the social costs of antisocial behavior, Twitter has another feature that many social media platforms lack: it allows ordinary people to directly contact powerful politicians and celebrities. This operates on the other side of the cost-benefit analysis, since one feels more powerful when one controls or manipulates an otherwise powerful individual.

*Jimmy Dean Sausage*: Hi Ken - We have a package waiting for you at FedEx in Pensacola, and they need your phone number. Are you able to post it in a private message so that you can receive it, or can you share it on here? Thank you!

*KenM*: THANK YOU i will have my grandson drive me to Fed Ex this evening after church, here is my #SSN: 457-55-5462.<sup>21</sup>

KenM's masterful work meets none of the conditions of the account, but it seems to be a paradigm instance of trolling. (In particular, it is a textbook case of trolling constituted by a sustained communicative interaction, rather than a single speech act.)

### 2.3 A Revised Account

To perspicuously revise the account of trolling in light of the KenM example we will need to introduce a term that does not, as far as I am aware, exist in popular discourse. I will use the phrase “trolling for” to specify the reaction that an instance of trolling is likely, believed, and intended to elicit. According to the account developed so far, all trolling is trolling for alarm or outrage. Of course, it may have troubled some readers to see this disjunction in the account in the first place. If one can troll for one of two related but distinct reactions, then we might wonder what unifies them and whether there are other reactions for which one can troll. This concern is supported by the fact that KenM quite clearly seems to be trolling for a very different kind of response—something like: *overly-earnest and polite corporate customer-service concern*. KenM is making fun of both (a) older people who suffer from internet illiteracy and (b) Jimmy Dean Sausage for its unflinching corporate pandering. But though the former group are among the objects of KenM's mockery, they are *not* the targets of his trolling. He is not trolling elderly internet users. He is trolling Jimmy Dean Sausage. And he is trolling not for outrage, but for a reaction that we might simply call “corporate concern.” His intention is for the representatives of Jimmy Dean Sausage—who begin every reply with a polite and exhaustingly chipper “Hi Ken”—to display the kind of patience and studied concern that characterize customer service representatives in the 21st century.

What are the reactions for which one can troll? One can troll for (almost) any reaction that one regards as mockworthy.<sup>22</sup> Here, then, is the final account of trolling, stated somewhat more schematically. A singular or group agent  $\mathcal{A}$  trolls an audience  $B$  by making an utterance or series of utterances that:

- (1) is likely to elicit from  $B$  a reaction  $r$  that  $\mathcal{A}$  regards as mockworthy, and
- (2) is believed by  $\mathcal{A}$  to be likely to elicit  $r$  from  $B$ , and

<sup>21</sup> This is the social security number of Todd Davis, who publicized it himself, not that originally used by KenM.

<sup>22</sup> The ‘almost’ is explained in section 3.1.

(3) is primarily intended by  $\mathcal{A}$  to elicit  $r$  from  $B$ .<sup>23</sup>

Contemporary internet vernacular labels this mockery with the expression “lulz,” which is a nefarious adaptation of “LOL”.<sup>24</sup> What it is, exactly, to take some reaction or behavior-type to be mockworthy is itself a non-trivial philosophical question. But it cannot be fully answered here. An account of trolling that makes use of a concept like *mockery*, which is perhaps as mysterious as *trolling*, may still be illuminating. The demand that no concept be left unexplained leads quickly to regress.

Without giving a full account, however, mockery seems to involve both amusement and negative assessment. Plausibly, the trolls mentioned so far have these attitudes toward the reactions that they intend to elicit. We can conjecture, to mention two examples, that Jenkem trolls take pearl-clutching panic over teenage drug trends to be laughably alarmist and that KenM takes corporate concern to be comically insincere.<sup>25</sup>

There is one potential worry about this account that is worth addressing immediately: the account might seem over-inclusive. Consider, for example, a situation where one sibling says “your dress is ugly,” knowing and intending that this statement will send the other sibling into a tantrum, which the first sibling finds mockworthy. We would not typically think of this as trolling, but it meets the conditions set out in the account.<sup>26</sup>

Although we might not instinctively and immediately label this behavior “trolling,” nonetheless, I would like to suggest that it is trolling. Indeed, it is quite natural to imagine a parent advising the upset child by saying, “Your sibling knows how sensitive you are about your clothes. She’s trolling you, just like people do on the internet. Try not to give her the satisfaction of a response, and she’ll stop.” To my ear, there is nothing unusual or suspicious about the use of “trolling” in this parental advice. And it is good advice. This sibling-dress-comment example and the various examples of trolling mentioned so far do all seem somewhat akin. And it counts in favor of the account of trolling presented here that it explains the way in which they are akin.

But the parental advice just mentioned sounds like the kind of advice typically given to children who are suffering from bullying. So it is reasonable to wonder: how exactly is trolling different from ordinary bullying? The account just presented provides an answer. First, it allows that some trolling

---

<sup>23</sup> Mockery does seem to have something central to do with trolling, but it might be wondered why the reaction is regarded as mockworthy rather than the act of trolling itself being a form of mockery. Aside from possible counterexamples, this won’t work because it seems that the reaction is the thing that would be mocked, but at the time of trolling that reaction has not yet occurred and so cannot be mocked.

<sup>24</sup> “LOL,” standing for laughing out loud, is one of the earliest elements of internet slang.

<sup>25</sup> A possible consequence (depending on what auxiliary hypotheses one is willing to take on) of the account presented in this section, is that not all trolling is morally bad, or even that the fact that some type of behavior is still is a moral reason against exhibiting it. If this is a consequence, I do not think it is a mistaken one.

<sup>26</sup> Thanks to an anonymous reviewer from *Public Affairs Quarterly* for raising this concern.

is also bullying. But, second, it yields the result that many instances of bullying—even including paradigm instances—are not also instances of trolling. According to the account of trolling presented here, trolling is intended to elicit a certain kind of reaction. But much bullying is not intended to elicit a particular reaction, or any reaction at all.<sup>27</sup> Consider the classic scene in which a child is holding her notepad and school books, and a bully walks by and slaps the top of the pile of books and paper, sending them scattering to the floor. Of course, the bully *may* intend to elicit some observable reaction. She may intend for her victim to run off crying, to launch into a didactic speech, or something else. In such a case, assuming the other conditions are met, the account presented here yields the result that this is trolling. But more typically, when a school bully slaps the books out of another student's hands, the bully is not aiming to elicit a particular reaction. Indeed, if the victim failed to react in any observable way, remaining silent and waiting until the bully has passed by to pick up the books, then most bullies would consider that a success. And in such cases, the account presented here would not classify this somewhat paradigm instance of bullying as trolling. This seems to me like the correct result. And it is the same result we get if the sibling-dress-comment example is modified along similar lines. If the sibling says “your dress is ugly,” not to provoke an outburst, but as part of a cruel and long-standing campaign to demoralize, in which each instance of behavior is met only with silence, then the sibling is a bully, but not a troll.

### 3 A Ubiquitous Feature

There is another feature of trolling—one that seems present in all or nearly all instances—that was left out of the above account because it seems to be less part of what trolling is, and more something that attaches to trolling as a *result* of what it is (perhaps in combination with some stable features of human nature).

#### 3.1 Targeting Sincere Concern

The feature is this: trolls attack those who have or present themselves as having sincere concern or non-ironic attachment to something. The troll intends to elicit a reaction that she regards as mockworthy, but not just any reaction will do. Some reactions express playful or ironic or indifferent attitudes. The troll does not seek to elicit these reactions. Rather, she has derision for, and therefore directs her trolling toward those who *care* about something. As Barney's Aristotle explains, all communities of speakers are potential targets of trolling because such communities are unified by shared sincere concern—every such community “holds certain goods in common.”<sup>28</sup>

---

<sup>27</sup> As I am understanding “reaction,” a lack of speech or movement does not constitute a reaction. This is, at least, one common way of using the word “reaction.” It is possible, on this use of the term, to truly say of someone that they “did not react.” Of course, there is an alternative understanding of the term on which even motionless silence constitutes a reaction. That is not how I am using the term.

<sup>28</sup> Barney 2016, 194.

The Jenkem trolls, for instance, targeted concern about teen drug use, which they perhaps regarded as overwrought hand-wringing. Another vivid example comes from Limond, who enjoys trolling atheists even though he is himself an atheist [*viz* throughout]:

I tweeted, “It's such a shame that athiests will never know true love. #atheism.” Look at the misspelling of “atheists” – what pedant could possibly resist that bait? Not many, and I must have had a dozen catches almost straight away. You know the type. Militants. Wanks. The type that have “atheist” in their bio, like anybody gives a fuck. The type that searches for #atheist on Twitter. And I argued with a handful of them simultaneously for close to an hour, giggling like a schoolboy as I typed, “Your going too hell!” They didn't know whether to go after the religious debate or the grammar. It was hilarious.<sup>29</sup>

Limond's laughter is directed at atheists who care about serious and sincere arguments for atheism, and about proper spelling and grammar.<sup>30</sup>

As a final example, it is worth mentioning RIP trolling—the trolling of online memorial pages. In 2006, a seventh grader with no history of depression, Mitchell Henderson, shot himself in the head with his parents' rifle. Some of Henderson's classmates created a memorial page on the then-popular social media site, MySpace. The memorial page was discovered by users of a famous trolling message board called “/b/” at 4chan.com, who noted that many of the posts on the page—and recall that Henderson's classmates were 12 years old—were ungrammatical. One of Henderson's classmates wrote that Henderson was “an hero to take that shot, to leave us all behind. God do we wish we could take it back...” Another posted a memory of Henderson having lost his iPod. The users of /b/ began joking that Henderson had killed himself because he lost his iPod and began using the phrase “an hero” as a verb, meaning to kill oneself.<sup>31</sup> One /b/-user placed an iPod on Henderson's grave in Rochester, Minnesota, took a picture, and posted it to /b/. According to the account presented here, none of these morbid jokes constitute trolling. To be sure, the denizens of /b/ were mocking a deceased 12-year-old boy and his mourning classmates, but they were not trolling those classmates. Indeed, at this point in the story, the classmates were unaware of any of the activity on /b/.

---

<sup>29</sup> Limond 2012.

<sup>30</sup> The KenM from earlier might seem like a counterexample because the attitude he attempted to elicit from Jimmy Dean Sausage—corporate concern—expresses *insincerity*. But things are not so simple. Jimmy Dean Sausage (or its representatives) are seriously concerned about something: appearing concerned. These corporate representatives may not care about their customers, but they care about looking like they care. And that concern is KenM's target.

<sup>31</sup> As in, ‘So-and-so is worthless. He should an hero himself.’

But then users of /b/ began posting on Henderson’s memorial page itself, including photos of Henderson edited to include iPods, dancing iPods, spinning iPods, and unrelated pornography. Then, as Henderson’s father recalls, the family began receiving phone calls: “It sounded like kids. They’d say, ‘Hi, this is Mitchell, I’m at the cemetery.’ ‘Hi, I’ve got Mitchell’s iPod.’ ‘Hi, I’m Mitchell’s ghost, the front door is locked. Can you come down and let me in?’ It really got to my wife.”<sup>32</sup> Henderson’s parents continued receiving these phone calls for a year and a half.

Many similar cases exist.<sup>33</sup> Without claiming to explain everything about this perplexing phenomenon, there are some illuminating things that can be said about RIP trolling. The first is that it is worth distinguishing (a) the posts made on the 4chan.com message board, (b) the posts made on the MySpace memorial page, and (c) the phone calls made to the Henderson household. As mentioned earlier, (a) are not instances of trolling. And it is not clear whether (c) are either. It is hard to say what reactions those who placed those phone calls were hoping to elicit, if any. And it is harder to see how they might be understood by anyone as mockworthy. However, I do not think this uncertainty is a problem for the account of trolling presented here. Such a problem would arise if we had a clear intuition that these phone calls to the parents of a recently deceased child were instances of trolling, but the account failed to categorize them as such. But I, for one, do not have the clear intuition that those phone calls are instances of trolling.

The posts to the MySpace memorial page, on the other hand, do seem to be clear instances of trolling. So the account ought to explain them—or, at minimum, it ought to classify them as instances of trolling. The motivation of those who posted various edited photos on the Henderson memorial page remain unclear. But understanding that trolls target sincere concern does seem to go some small part of the way toward explaining their mindset.<sup>34</sup> As far as the Internet goes, it does not get much more sincere and earnest—or, at minimum, one cannot find more of a *presentation* of sincerity and earnestness—than a memorial page.

Those familiar with RIP trolling are aware that even before the trolls arrive not all posts on internet memorial pages are straightforwardly sincere. There are *grief tourists*, who Whitney Phillips, a researcher specializing in internet trolling, describes as “people who comment on the pages of dead

---

<sup>32</sup> Schwartz 2008.

<sup>33</sup> This includes the trolling of the Facebook memorial page of Alexis Pilkington, a high school girl who died by suicide and was mocked as a “suicidal slut,” and the trolling of the family of Nikki Catsouras, an 18-year-old who fatally crashed her father’s car and whose parents were then sent leaked pictures of her mutilated corpse.

<sup>34</sup> As Hess (2017) puts it: “Internet trolls work by exploiting the gap between the virtual and the real. They float, weightless and anonymous, across the web, then reach out and rattle people who are pinned down by fixed ideologies, moral codes and human emotions. Any attachment to principles — even really basic ones like ‘don’t torture grieving parents’ — gives the troll an opening. [Trolling has always been] about the distance between people who care and people who don’t.”

strangers in order to mourn alongside other strangers.”<sup>35</sup> Grief tourists, it might be argued, do not have sincere concern for the deceased. This, it seems to me, is a potentially contentious claim, since it is hard to know exactly what attitude grief tourists have toward those deceased people who they have come to mourn, even though they have never met those people. But let us assume that they do not have sincere concern for the deceased. Still, two things about grief tourists are worth noting. First, although they might not have sincere concern for the deceased, they very plausibly have sincere concern for the seriousness and sanctity of memorial pages. After all, their interest in mourning for and alongside strangers is only satisfied if they and the relevant memorial page remains solemn and undisturbed by RIP trolls. Second, although grief tourists might not have sincere concern for the deceased, they do, by definition, *present themselves* as having that sincere concern. So it is still accurate to say of RIP trolls whose primary targets include grief tourists that they attack those who have or present themselves as having sincere concern or non-ironic attachment to something.

But there is another worry about RIP trolls that is worth addressing. They might not fit into the official account of trolling presented in the previous section, because it is not clear that RIP trolls target a reaction that they regard as mockworthy. In responding to this worry, it should be noted that RIP trolls need not find expressions of mourning to be mockworthy (though they may). Expressions of mourning in online memorial pages are not the relevant *reactions* to trolling behavior. Rather, what we find in cases of RIP trolling is a memorial page where expressions of mourning are being posted. RIP trolls intrude and cause a variety of reactions. These reactions are not themselves more mourning. Instead they are, for example expressions of shock and outrage at the fact that a place reserved for solemn and sincere expressions has been intruded on in such a way. The question, then, is whether it is plausible that RIP trolls find these reactions mockworthy. And Phillips’s research seems to suggest that they do.<sup>36</sup> A consistent finding from Phillips’s research is that RIP trolls find it laughable that anyone would be shocked and outraged by the misuse of a place reserved for solemn and sincere expressions because they think that those places are already full of insincere expressions—i.e. grief tourism. And moreover, even if we put grief tourism entirely aside, there is another reaction that many RIP trolls seem to be trolling for: condemnatory and disapproving news media coverage. Phillips notes that many RIP trolls find this kind of media reaction laughably sanctimonious precisely because of “just how good personal tragedy is for the business of media.”<sup>37</sup> And once we recognize that that kind of coverage is often the reaction intended by RIP trolls, it becomes much less mysterious how that reaction might be regarded as mockworthy.

---

<sup>35</sup> Phillips 2015a, 24.

<sup>36</sup> Phillips 2015a, 71–94. The place within Phillips’s book to focus for the most significant discussion of RIP trolling is Chapter 5, entitled “LOLing at Tragedy: Facebook Trolls, Memorial Pages, and the Business of Mass-Mediated Disaster Narratives.”

<sup>37</sup> Phillips 2015a, 87.

### 3.2 The Troll's Own Sincere Concern

But while on the topic of sincere concern, it would be negligent not to note that trolls are often motivated by very sincere concerns of their own. This is not a contradiction. The troll does not believe two inconsistent propositions. Nor does this, I think, even rise to the level of hypocrisy. But it is important. The troll can, so to speak, dish it out, but cannot take it. Or, similarly, the troll is open to the same kind of attack that she levels on others.<sup>38</sup>

A prominent internet troll, who was imprisoned for his behavior and operates under the pseudonym “Paulie Socash,” was interviewed by a social scientist who studies trolling and said the following about RIP trolling:

People are shocked that people could be so low as to say mean things about dead kids to their families or whatever. The reality of this is simple: the vast majority of those who get large memorial pages on Facebook are cute little kids (Jamie Bulger) or pretty young ladies (Jenni-Lyn Watson, Chelsea King) or useful pawns for a cause (Tyler Clementi and other gay suicides). These memorial pages are decidedly not a place for friends and family to grieve (family and friends should be grieving together in private like normal people). In reality, these are havens for ‘grief tourists’: people who substitute online emotions and declarations of solidarity for real emotional relationships and friendship. Most memorial pages are not set up by friends or family; they are created by people who are too involved with the stories they read online or see on the news - people who derive some sense of self-importance and worth from being seen to care by strangers.<sup>39</sup>

If Socash’s attitude is shared by other trolls, then they appear to be motivated by their own sincere concern regarding “grief tourism.” Socash himself notes the tension: “It is, after all, earnestness and selfrighteousness that are the best things to attack when trolling, so having set positions of one’s own is a problem.”<sup>40</sup> Socash is right that it is a problem, but it is not so much a *philosophical* problem as much as it is a *practical* problem *for the troll*.

---

<sup>38</sup> It seems *prima facie* plausible that most trolls are men (and empirical evidence supports this (Ferenczi, Marshall, and Bejanyan 2017)), though there are many notable exceptions, such as the Israeli comedian, Racheli Rottner. After giving considerable thought to the issue, I have decided, however, to continue my standard practice of using feminine pronouns when the gender of the antecedent is unknown.

<sup>39</sup> Phillips 2011, 75. This is from an interview conducted by Whitney Phillips, Assistant Professor of Communication, Culture, and Digital Technologies at Syracuse University, and whose 2015a book, *This is Why We Can't Have Nice Things*, is an invaluable resource on trolling and other internet media phenomena.

<sup>40</sup> Phillips 2011, 74.

And the problem gets worse. Not only are trolls often motivated by some general sincere concern, but because trolling is itself a purposeful, goal-oriented activity, it *necessarily* involves a more specific sincere concern: concern for eliciting the reaction for which one is trolling.<sup>41</sup> The result is that all trolls are themselves *potentially* susceptible to trolling. Given the account of trolling introduced so far, what is also needed to be susceptible to trolling is for this sincere concern to manifest itself as a reaction to some-or-other kind of social interaction. As discussed in section 5, therefore, one way to respond to a troll—at least one whose sincere concern manifests itself in this way—is to figure out what she cares about, and troll her back.

#### 4 Non-Essential Features of Trolling

There are several other features of trolling that have been thought to be central to the phenomenon. Though the account presented here does not take them to be essential or necessary, they may nonetheless be paradigmatic. Very briefly, here are four such features.

First, the troll's primary intention is to elicit some kind of putatively mockworthy reaction, but the troll often has a certain secondary intention, to be achieved by way of achieving the primary intention: to disrupt a discursive community or an ongoing conversation. For instance, a troll might say inflammatory things within a community so as to elicit outrage, disrupt the community, and distract it from more productive discourse. As Barney's Aristotle puts it, a troll:

...speaks to a community and as being part of the community; only he is not part of it, but opposed. And the community has some good in common, and this the troll must know, and what things prompt and destroy it: for he seeks to destroy.<sup>42</sup>

Others have also remarked on this feature of trolling as well.<sup>43</sup> But KenM's behavior seems to be a central instance of trolling that does not involve disrupting a community. Perhaps some corporate Facebook pages receive consistent interaction from the same group of internet users, who know one another, have ongoing conversations, and have built a community of sorts. But, undoubtedly, some

---

<sup>41</sup> In a Kantian voice, susceptibility to the possibility of trolling is a condition of the possibility of trolling.

<sup>42</sup> Barney 2016, 193.

<sup>43</sup> It is often thought that after infiltrating a group trolls “set about disrupting proceedings while trying to maintain cover” (Dahlberg 2001). Similarly, “...the peculiarity [*idion*] of the troll is not annoyance or controversy in general, but confusion and strife among a community who really agree.” (Barney 2016, 194). Claire Hardaker, a linguist, has conducted a study of internet trolling behavior in which she defines a troll as follows: “...this paper suggests the following, working definition of the term “troll”: A troller is a CMC [computer-mediated communication] user who constructs the identity of sincerely wishing to be part of the group in question, including professing, or conveying pseudo-sincere intentions, but whose real intention(s) is/are to cause disruption and/or to trigger or exacerbate conflict for the purposes of their own amusement.” (Hardaker 2010, 237).

do not. Assume that the Facebook page of Jimmy Dean Sausage is of the latter kind. Still, even with that assumption, KenM's sustained interaction on that page is a central example of trolling. Of course, we could expand the notion of a *community* such that even an infrequently trafficked Jimmy Dean Facebook page qualifies, and under that understanding we could include disruption of a community as a necessary feature in our account of trolling. But we will have then defined "community" so broadly as to have made the fact that trolls aim to disrupt such communities far less informative.

A second paradigm feature of trolling is noted by Phillips, who discusses the exact same Jenkem incident mentioned earlier, and points out that:

...the trolls knew exactly how to manipulate the news cycle, and in the process forwarded an implicit critique of the ways in which media research and report the news. Specifically, many outlets are so eager to present the latest, weirdest, and most sensational story that producers often fail to conduct even the most cursory background research—or worse, they conduct the appropriate background research, but choose to run misleading segments anyway.<sup>44</sup>

A good deal of trolling can be understood this way—as a means of indirectly critiquing news media or some other cultural system. The Jenkem ruse fits this description perfectly. And even KenM's behavior can be seen as exhibiting a feature of this sort, so long as we expand the target of critique to include social media and corporate culture in general, and not merely professional news media, which Phillips has in mind in the above passage. But there are many other examples—such as Limond's behavior on Xbox Live or Gubin's "Welcome to Cleveland" sign—that seem like more-or-less central instances of trolling, but which cannot comfortably be understood as implicit or explicit critiques of anything, including traditional news media, social media, or any other identifiable cultural systems. So implicitly expressing this kind of critique seems to be an important, but non-necessary feature of trolling.

Third, trolling often involves *non-committal speech*. That is, it often involves speaking playfully or in jest, in such a way that the speaker is not normatively committed, as one is by normal speech acts like assertions, questions, promises, etc. Without saying how to understand this type of speech—which can be understood as the suspension of the conversational maxim of quality, as using words as props in a game of make believe, as not updating the conversational common ground or the score in the language game, as updating a pretend common ground or score in the language game, etc.—it is clear that it is an expansive and diverse category that includes: acting in a play,

---

<sup>44</sup> Phillips 2015a, 6.

writing fiction, and many forms of joking.<sup>45</sup> Jean-Paul Sartre points out that anti-Semites often use non-committal speech, and the point here is that trolls do the same.<sup>46</sup> The troll speaks frivolously, and when pressed will often claim that she was only joking.<sup>47</sup>

But there is a related, fourth feature of trolling that is not essential, but very much seems characteristic of the phenomenon. When trolls use non-committal speech, they often do so in a particular disingenuous way. They intend that their non-committal speech is taken, at least momentarily, to be committal.<sup>48</sup> For instance, when Gubin paints “Welcome to Cleveland” in enormous letters on the roof of his art studio near the Milwaukee airport, he intends for airline passengers to, at least momentarily, take him seriously.<sup>49</sup> Gubin is joking. He is not welcoming anyone to Cleveland, and he knows it. But the joke only works if his audience is not in on it.<sup>50</sup>

This final non-essential feature of trolling—the disingenuous use of non-committal speech—is perhaps the most illuminating of the three. First, it partly explains why trolling, which can sometimes be benign, can be devastating. When a troll purposefully presents non-committal speech as committal, it can function as a form of gaslighting—potentially leaving the audience unable to determine whether they can accurately distinguish committal and non-committal speech. Second, the disingenuous use of non-committal speech suggests something about the moral status of trolling

---

<sup>45</sup> The approximate consensus (*pace* Currie 1986) is that fiction, or fictional speech, should be understood broadly in terms of pretending to perform speech acts and not in terms of performing some distinct type of speech act that is properly listed alongside assertion and so on. The reason that there is no distinctive fictional illocutionary force is that this kind of speech does not involve bare propositional (or more broadly sentential) contents. For an account of fictional discourse as involving the suspension of something like a Gricean maxim of quality, see A. Martinich 2001; A. P. Martinich 1980. For a similar discussion relating to humor, see Raskin 1984. For an account in which words are props in a game of make believe, see Walton 1990; Currie 1990. Walton explicitly rejects approaches to fiction that characterize them in terms of speech acts, but it has been plausibly argued that his own theory can be cast in those terms (see Hoffman 2004). For work on scorekeeping and conversational common ground, see Brandom 1994; R. C. Stalnaker 1970; R. Stalnaker 1973, 1974; MacFarlane 2011; C. Roberts 2004; Craige Roberts 2012. For an account of fictional discourse in terms of non-deceptive pretend assertion, see Searle 1979. Meinongians, like Parsons (1980), will not accept any kind of pretend-based account of fictional speech.

<sup>46</sup> Sartre 1948, 13–14. I discuss the relevant passages from *Anti-Semite and Jew* elsewhere, but do not discuss them at length here so as to conserve space.

<sup>47</sup> As Sartre (1948, 14) might put it, she “loftily indicat[es] by some phrase that the time for argument is past.” It should be noted that not all mockery is non-committal. For instance, one might (non-trollingly) mock, saying “You are ugly, ha ha!” and, when pressed by the object of her insult, she might insist that although she was mocking her, she was serious about it: “Yes, I meant it. I genuinely think you are ugly and it is funny to me that you have this misfortune.”

<sup>48</sup> Whether this kind of disingenuous use of non-committal speech is simply a form of insincere speech is a matter that is beyond the scope of this paper and also beyond the scope of my own knowledge.

<sup>49</sup> Or, at minimum, he intends for them not to know whether to take him seriously or not.

<sup>50</sup> Two things are worth noting about this particular brand of disingenuous non-committal speech that is characteristic, though not essential, of trolling. First, this is not non-committal speech that is intended to be taken as such but is mistakenly taken as committal—that is, so to speak, we are not talking about jokes that bomb. Second, this is not committal speech which is responded to negatively leading the speaker to disingenuously suggest that the speech was non-committal all along when in fact it was committal—that is, we are not talking about when someone says something seriously and then after criticism pretends that she was just joking all along.

(though that is not the topic of the present paper). There is nothing morally bad about non-committal speech *per se*. Joking is great. But the disingenuous non-committal speech that involves deception is a good candidate for an immoral feature of trolling.<sup>51</sup>

## 5 Practical Upshot

It cannot be overemphasized that the following section does not offer *advice*. Philosophers are simply not qualified to offer practical guidance. That being said, having an account of the nature of trolling is helpful when formulating a response. For instance, consider the following advice that Robert Paul Wolff offered to students at the University of California, Berkeley in regards to a right-wing troll who was set to speak on campus:

Permit the speaker to speak, but either refuse to attend, or attend and sit in stony silence while he or she speaks.

My favorite solution is to mob the site, filling every available seat, and then just sit. No boos, no catcalls, no demands for equal time. Just sit. Let the person speak for as long as he or she wishes, but just sit. Trust me, this would be unnerving. My guess is that someone like Yiannopoulos would start out bold and brazen, making deliberately inflammatory statements to evoke some response, and then begin to falter as the minutes go by and he gets no response at all. Try it some time. After a while, when it turns out that he is getting the silent treatment, he will make a series of abusive statements and then crawl away.

Suppose there are a few supporters in the audience. Fine, let them cheer and applaud, to stony silence from everyone else. This is what is often called “shunning,” and the psychological effects can be quite forceful.

The freedom to speak does not carry with it a right to be responded to, or even listened to.<sup>52</sup>

---

<sup>51</sup> The other obvious point about the moral status of trolling relates to how directly intending to elicit a reaction from one’s interlocutor, rather than attempting to rationally persuade them, has an affinity with the Kantian wrong-making feature of treating others as means rather than ends, or, similarly, influencing others in a way that they would themselves reject, or that one would not be willing to generalize. (See Parfit 2011, chap. 9). Also, as Thomas Hill puts it, “The idea that one should try to reason with others rather than to manipulate them by nonrational techniques is manifest in Kant’s discussion of the duty to respect others.” (Hill 1980, 96).

<sup>52</sup> Wolff 2017. It should be noted that Wolff’s suggestion of how Yiannopoulos is likely to react to the silent treatment is not dissimilar to Sartre’s characterization of how the anti-Semite is likely to react to certain kinds of engagement (Sartre 1948).

Is Wolff offering good advice? I do not know.<sup>53</sup> But it does fit with the nature of trolling, as presented here. If one's goal is to discourage trolling (and I do not assume that it always is), then the silence that Wolff recommends may have this effect.<sup>54</sup>

There are, however, alternatives to silence. The alternative mentioned in section 3—trolling the troll back—comes in at least two varieties: (a) targeting some contingent sincere concern that the troll happens to have or (b) targeting the sincere concern for trolling itself that the troll necessarily has. As an example of the former, trolls often have a sincere concern for free speech. Something simple and purposefully ill-considered—e.g., “Actually, it was controversial when they said it, but many scientists have shown that free speech is rarely needed.”—may do the trick. If the troll sputters back something like, “How could ‘scientists’ have possibly shown that?! And if it was controversial, then they needed free speech to say it in the first place!” then one has successfully turned the tables.

To understand the second variety—targeting the troll's necessary sincere concern for trolling itself—recall the Jenkem example with which we began.<sup>55</sup> But imagine, counterfactually, that a news outlet recognized Jenkem as an act of trolling and wished to troll the trolls back. One way to do this would be to appear credulous while repeatedly asking for more details and evidence, stringing the trolls along for weeks or months with no intention of ever airing or publishing a report on the supposed drug trend. Of course, on the account presented here, their own behavior will only count as trolling if they regard the trolls repeated efforts of mockworthy.<sup>56</sup>

Perhaps obviously, this approach—since it is different from silence and has a different aim as well—might lead to *more* trolling, not less. But it is sensitive to the nature of trolling. And, again, the purpose of all of this is not to give advice, but only to gesture at the usefulness of the kinds of metaphysical account of a social phenomenon that this paper offers.

---

<sup>53</sup> It is also difficult advice to implement. It is not easy to ignore something specifically designed to be provocative, and it can be near impossible to coordinate a large group of undisciplined and passionate activists to exhibit this kind of restraint. It is particularly difficult to ignore a troll because trolling is often designed to attack insecurity. See Schwartz 2008. Moreover, a silent response will be impossible to implement if those responding are interested in moral grandstanding. See Tosi and Warmke 2016.

<sup>54</sup> As mentioned in section 2, trolling is a way of exercising power over others. Since the value of trolling to the troll is instrumental in this way, it is not hard to see how to remove that value and undermine the troll's motivation. If, for instance, the troll says offensive things in order to provoke outrage, then refusing to be outraged will deplete trolling of its appeal.

<sup>55</sup> Phillips 2015a, [b] 2015.

<sup>56</sup> Though the troll always has sincere concern for the trolling in which she engages, it might still be that she is engaging in a form of trolling that attempts to elicit trolling as a response. The troll might intend to be revealed as a troll and trolled back. I am not aware of the actual existence of such a case of higher-order trolling, but I see no reason why it would be impossible.

And this kind of account fits with another potential response—and I am *certainly* not recommending this one: simply mocking the troll. Here is a comment posted on an online forum in which a troll had been posting (the example is preserved by Hardaker 2010, 235):

Do you think [troll’s username] made up her hotmail address all special for us, just for this post? (Google is your friend.) Do you think she really has a husband? Do you think she really even is a \_she\_?

Here the troll is being mocked for taking the time to create an entirely new email address just to anonymously post one message. That fact is a potential target for mockery, and it is such a target precisely because the troll has the second type of sincere concern introduced in section 3.1.

What these ways of responding to a troll have in common is that they are all specifications of the often-heard but less-often-understood advice: “don’t feed the troll.” It is natural to hear that mantra and take it to be recommending silence, such as the kind recommended by Wolff. But perhaps one of the primary advantages of having an account of trolling, like the one presented here, is that it brings forth other, non-silence responses that avoid feeding the troll. Once we see that trolls are attempting to elicit some kind of response, it becomes clear that many other responses can starve them as well.<sup>57</sup>

University of North Carolina at Greensboro

Thanks to Adam Bradley, Ravit Dotan, Hannah Ginsborg, Aaron James, Niko Kolodny, and Nick Riggle for reading and commenting on early writings that would eventually become this paper.

---

<sup>57</sup> How, though, does one know which type of response a particular troll intends to elicit? The answer is that one must make an educated guess. But this is the same kind of guess that is involved in all human interaction. To deftly get by in the world, we have to figure out what other people are thinking.

## References

- Barney, Rachel. 2016. “[Aristotle], On Trolling.” *Journal of the American Philosophical Association* 2 (2). Cambridge University Press: 193–95.
- Boudourides, Moses A. 1995. “Social and Psychological Effects in Computer-Mediated Communication.” Retrieved February 26: 2005.
- Brandom, Robert B. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
- Cheng, Justin, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. “Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions.” *CSCW : Proceedings of the Conference on Computer-Supported Cooperative Work. Conference on Computer-Supported Cooperative Work 2017* (February): 1217–30.
- Coons, Christian, and Michael Weber. 2014. *Manipulation: Theory and Practice*. Oxford University Press.
- Currie, Gregory. 1986. “Works of Fiction and Illocutionary Acts.” *Philosophy and Literature* 10 (2). Johns Hopkins University Press: 304–8.
- . 1990. *The Nature of Fiction* Cambridge University Press.
- Dahlberg, Lincoln. 2001. “The Internet and Democratic Discourse: Exploring The Prospects of Online Deliberative Forums Extending the Public Sphere.” *Information, Communication and Society* 4 (4). Routledge: 615–33.
- Douglas, Karen M., and Craig McGarty. 2001. “Identifiability and Self-Presentation: Computer-Mediated Communication and Intergroup Interaction.” *British Journal of Social Psychology* 40 (3). Wiley Online Library: 399–416.
- Ferenczi, Nelli, Tara C. Marshall, and Kathrine Bejanyan. 2017. “Are Sex Differences in Antisocial and Prosocial Facebook Use Explained by Narcissism and Relational Self-Construal?” *Computers in Human Behavior* 77 (December): 25–31.
- Hardaker, Claire. 2010. “Trolling in Asynchronous Computer-Mediated Communication: From User Discussions to Academic Definitions.” *Journal of Politeness Research. Language, Behaviour, Culture* 6 (2): 277.
- Hess, Amanda. 2017. “How the Trolls Stole Washington.” *NYTimes Magazine*, February.
- Hill, Thomas E. 1980. “Review: Benton, Kant’s Second ‘Critique’ and the Problem of ‘Transcendental Arguments.’” *Journal of the History of Philosophy* 18 (3). Johns Hopkins University Press: 356–57.
- Hoffman, Sarah. 2004. “Fiction as Action.” *Philosophia* 31 (3): 513–29.
- Hume, David. 2006. *An Enquiry Concerning the Principles of Morals: A Critical Edition* Clarendon Press.
- Kiesler, Sara, Jane Siegel, and Timothy W. McGuire. 1984. “Social Psychological Aspects of Computer-Mediated Communication.” *The American Psychologist* 39 (10). American Psychological Association: 1123.
- Limond, Brian. 2012. “Limmy’s Show: Confessions of an Internet Troll.” *The Guardian*, November 9.
- MacCormick, Neil. 1983. “What Is Wrong with Deceit?” *The Sydney Law Review* 10: 5–19.
- MacFarlane, John. 2011. “What Is Assertion?” In *Assertion*, edited by Jessica Brown and Herman Cappelen. Oxford University Press.
- Mannison, D. S. 1969. “Lying and Lies.” *Australasian Journal of Philosophy* 47 (2). Routledge: 132–44.
- Martinich, Aloysius. 2001. “A Theory of Fiction.” *Philosophy and Literature* 25 (1). Johns Hopkins University Press: 96–112.
- Martinich, A. P. 1980. “Conversational Maxims and Some Philosophical Problems.” *The Philosophical Quarterly* 30 (120). Wiley-Blackwell: 215–28.
- Parfit, Derek. 2011. *On What Matters: Two-Volume Set*. OUP Oxford.
- Parsons, Terence. 1980. *Nonexistent Objects*. Yale University Press.

- Phillips, Whitney. 2011. "Meet the Trolls." *Index on Censorship* 40 (2). SAGE Publications Ltd: 68–76.
- . 2015a. *This Is Why We Can't Have Nice Things*. MIT Press.
- . 2015b. "Internet Troll Sub-Culture's Savage Spoofing of Mainstream Media [Excerpt]: A New Book Details the Success of Trolls in Turning the Media against Itself." *Scientific American*, May 15.
- Raskin, Victor. 1984. *Semantic Mechanisms of Humor* Reidel.
- Roberts, C. 2004. "Context in Dynamic Interpretation." In *The Handbook of Pragmatics* edited by Horn and Ward, 197–220. Routledge.
- Roberts, Craige. 2012. "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics." *Semantics and Pragmatics* 5: 1–69.
- Saar, John. 1972. "Carrier War." *Life Magazine*, February 4.
- Sartre, Jean-Paul. 1948. *Anti-Semite and Jew*. Translated by George J. Becker. Schocken Books.
- Schwartz, Mattathias. 2008. "The Trolls Among Us." *NYTimes Magazine*, August 3.
- Searle, John. 1979. "The Logical Status of Fictional Discourse." In *Expression and Meaning: Studies in the Theory of Speech Acts* 58–75.
- Stalnaker, Robert. 1973. "Presuppositions." *Journal of Philosophical Logic* 2 (4): 447–57.
- . 1974. "Pragmatic Presuppositions." In *Context and Content*, edited by Robert Stalnaker, 47–62. Oxford University Press.
- Stalnaker, Robert C. 1970. "Pragmatics." *Synthese* 22 (1-2). Springer: 272–89.
- Suler, John. 2004. "The Online Disinhibition Effect." *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society* 7 (3): 321–26.
- Tosi, Justin, and Brandon Warmke. 2016. "Moral Grandstanding." *Philosophy & Public Affairs* 44 (3): 197–217.
- Walton, Kendall L. 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts* Harvard University Press.
- Wolff, Robert Paul. 2017. "The Philosopher's Stone: A Commentary on the Passing Scene by Robert Paul Wolff." *A SUGGESTION FOR BERKELEY STUDENTS*. February 2.  
<http://robertpaulwolff.blogspot.com/2017/02/a-suggestion-for-berkeley-students.html>.