Melissa Kwan

Prof. Emilio Mora, TF Ryan Sirk

PHIL 11: Philosophy of Law

11 December 2019

Limitations of Mill's Justification of Free Expression

The free expression debate can be framed as a tradeoff between liberty and harm prevention. In *On Liberty*, Mill argues that liberty almost always prevails. He argues that free expression conduces to truth and autonomy: truth will emerge and remain justified through the testing of ideas, and individuals can achieve higher levels of utility when they reason for themselves. He argues that these considerations outweigh the *harm principle*, which states that ``the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection" (13). He thus uses the virtues of liberty to put expression in a class of exceptions to the harm principle, acts that can be protected even if they harm others. In this essay, I will examine the relationship between the expression of false information and Mill's ideals of truth and autonomy, and use the analysis to criticize the idealism of Mill's strong free expression protections. **I will argue that these strong protections fail to achieve the full extent of truth and autonomy that Mill suggests due to the limitations of our cognitive capacities. When factoring these limitations into Mill's framework, we can derive a weaker defense of free expression.**

Mill's Liberty of Thought and Discussion principle states that speech and expression belong to the class of exceptions to the harm principle, which means that we should almost always protect expression even if it does cause harm (15). The Liberty of Thought and Discussion principle justifies this exemption on the grounds that it conduces to truth and

autonomy. On utilitarian grounds, Mill argues that truth is most likely to emerge in a "marketplace of ideas" (21). No idea can be considered "infallible": through this process of discourse, we can challenge ideas and achieve intellectual progress (23). Moreover, Mill argues that free expression allows truth to flourish as "living truths" rather than "dead dogma," which leads to the kind of complacency that leads societies to stagnate (34). Finally, Mill argues that decisions reached autonomously lead to the highest levels of utility (14). He uses truth and autonomy to argue that free expression should be exempt from the harm principle in almost all cases.

By arguing that having the strongest free expression protections leads to truth and autonomy, Mill fails to account for the limitations of our cognitive capacities. In this essay, I will analyze two of these limitations:

1. Our cognitive resources are *finite*. It takes time and cognitive resources to investigate ideas, and it is impossible to thoroughly vet all of them. This scarcity of resources means that time spent defending a belief against blatant falsehoods is time not spent pursuing further unknown truths.
2. Our judgments are subject to *confirmation bias*, defined as the "tendency to interpret new evidence as confirmation of one's existing beliefs or theories." Our desire to be correct often corrupts our rational abilities, which means that

Given these limitations, I will argue that Mill's strong free expression protections fail to achieve the extent of truth and autonomy that he needs to support his argument that liberty almost always outweighs harm prevention. I will focus on fake news, which does not clearly violate the harm principle in a way that libel or shouting "fire" in a crowded theater might.[1]

[1] Throughout this essay, I will use the example of fake news to provide concrete examples of how unfettered speech can detract from truth and autonomy. The example is useful because it is does not clearly violate the harm principle

Disinformation can hinder our ability to discover the truth due to our finite cognitive resources. Mill does not account for this limitation, as he does not consider the verification of ideas to be costly. Moreover, in his "marketplace of ideas," ideas that are expressed are able to be evaluated freely. In reality, ideas are less free-flowing, and different sub-communities can look to different localized sources of truth without earnestly considering other sources. Thus, if we allow free speech protections for falsehoods like fake news, it is possible that the free expression can lead some people to land accidentally in a localized pocket of disinformation. (In these cases, people have the freedom to check their information against other sources, but I will later argue that the benefits of this freedom are undermined by confirmation bias.) At best, protecting disinformation on the grounds of free expression encourages a healthy skepticism. At worst, this protection causes outright deception. Nevertheless, protecting fake news forces people to spend energy critically evaluating sources that could have been spent elsewhere. These protections put a burden on our limited cognitive capacities.

Even if we are not being deceived, Mill's strong free expression protections still put a burden on our cognitive resources. Mill does not argue that we need to respect all ideas equally; he simply believes that false beliefs should not be restricted. However, the proliferation of false ideas *is* a problem when people are not aware of the truth, and are being deprived of the information necessary to make informed decisions for themselves. While we might accept a truth, we must recognize that others may be misinformed. Indeed, it may even be in others' best interests to reject that truth for their own personal gain. Thus, even if we accept the truth on our end, that does not give us a free pass to dismiss the objectors: if we care about the truth, we must devote time to countering opposing views and correcting the narrative. While countering

---

in a way that libel or shouting "fire" in a crowded theater might. Government censorship of the press has had a dangerous history, which makes it difficult to imagine government censorship of news, even if it is blatantly false.

opposing views helps us prevent the truth from becoming "dogma," this defense seems unnecessary when the truth in question is simply a fact. For example, suppose we are forced to defend a true news report against thousands of fake ones. Such a defense should be unwarranted — the qualifications of the news source and the witnesses should speak for themselves. However, if we protect this sort of free expression, these defenses become crucial. They will become too costly to the point where we are too focused on defending past truths than on pursuing further truths based on our foundations. In order to make progress in science, in political thought, or in other disciplines, we need to build on past results. While it is important to continually test past results, the marketplace of ideas that Mill describes would allow people to focus on undermining past conclusions instead of the more utilitarian purpose of using those ideas to further future progress. The definition of "truth" concerns not just ideas we already have been exposed to, but further truths that build upon our first conclusions. Thus, if we are too focused on refuting conspiracy theories, we are limited in the amount of forward progress we can make. If we spend too many resources bolstering our foundations, our pursuits of new truths will suffer.

By hindering our collective understanding of truths, unfettered expression can also undermine our autonomy. This result depends on our definition of autonomy. If we define autonomy as the absolute number of options one has, allowing the most free expression would certainly maximize it. However, autonomy is better understood as being able to form and revise your conception of your own good. This definition is supported on Mill's classic utilitarian grounds, even if we account for the special value Mill places on individual freedom. Mill would argue that beliefs obtained autonomously have more utility than beliefs dictated by some other power, even if the latter strictly leads to "better outcome" for the agent in question. However, he

makes this comparison in the extreme, in the case of total autonomy versus total dependence on others. By examining a subtler case, we can see that autonomy should account for the quality of options as well as the number of options. Suppose person A is given ten reports of an event, all of which happen to be false. Suppose person B is given three reports of the same event, one true and the other two false. Person A's situation represents a society with no restrictions on disinformation, and person B's situation represents a society with partial restrictions on disinformation. Both A and B had multiple choices in deciding which report they decided to accept; however, B was the only one who had access to the truth. Thus, person B would have more autonomy in her ability to form and revise her conception of her own good, even though A had strictly more options to choose from. Person B ultimately had a higher expected utility than Person A on a more limited choice of options. In a society with strong protections of these false reports, person A's chances of getting *all* false reports might be relatively high. (And if so, even if she encounters a true report in the future, confirmation bias might lead her to discount it.) Thus, we can justify our definition of autonomy on utilitarian terms.

On this definition, disinformation can lead to a decrease in autonomy for individuals. While truth may be desirable for a population overall, in many cases, it is advantageous for an individual person to spread falsehoods without regard to what is actually the case to exploit others for their personal gain. For example, consider the case where fake news and incendiary headlines are passed off as reality, and the creators of the disinformation discount the value of the credible news sources. The spread of these falsehoods might hinder people's ability to autonomously decide what is true and good for themselves. Mill argues that autonomous individuals have the highest level of utility, but autonomy is at its best when individuals have the space to decide where to devote their energy. When we defend campaigns of disinformation

under protections of free expression, we might be burdening people's capacity for deliberations that matter to them.

Previously, my examples focused on disinformation that detracts from true information. My points have generally rested on the cases where the "crowding out" effect means that people's exposure to false beliefs crowds out the space for true beliefs. Note that this crowding out stemmed from our limited cognitive capacities. I will now consider the effects of the second flaw in our reasoning: confirmation bias. Confirmation bias makes it more difficult for truth to emerge, even if we allow for awareness of all sides of an issue. Once our beliefs have been solidified, we will be less likely to accept alternatives. Thus, if our first impression is a falsehood and we later find out the truth, we still might discount it. I will use this to extend my argument that an unfettered "marketplace of ideas" can still hinder truth and autonomy even if a person has access to both true and false beliefs.

Suppose there are two groups, one which has been indoctrinated by a true belief and another which has been indoctrinated by a false belief. Mill's strong free speech protections may allow the unfettered spread of these ideas, but they cannot guarantee that there will be good-faith disagreement on either side. When this good-faith disagreement between parties does not exist, it becomes more difficult for truth to emerge. What Mill idealized as "discourse" might devolve into a tribal "us vs. them" mentality, where neither side truly makes an effort to understand the other. While echo chambers may form even without fake news, the problem is exacerbated with fake news because the disagreements become about basic facts instead of about ideologies. When groups disagree not only about arguments, but about fundamental premises, it becomes increasingly difficult to have productive disagreement in the mode that Mill idealizes. Exposure to a different set of core beliefs may lead you to find even more flawed ways to reinforce your

own — not out of desire for truth, but desire to be right. Confirmation bias suggests that exposure to both truths and falsehoods is not necessarily conducive to truth if we have already been indoctrinated by a falsehood.

When we account for our cognitive limitations, free expression overall leads to less truth and less autonomy than Mill suggests. Since truth and autonomy were used to justify the strength of the Liberty of Thought and Discussion principle, the principle should be weakened in light of these limitations. However, the considerations of the harm principle remain as strong as ever. Thus, my argument entails a weakening of Mill's proposed free speech protections, not a ban on free speech protections entirely. I am not arguing that Mill's framework is wrong; I am simply arguing that the weights of the liberty and the harm principles should be different. In particular, autonomy and truth should be less powerful, because they do not follow as directly from strong free speech protections as Mill thought they did. This makes the harm principle comparably more powerful, which means that fewer expressions are exempt than Mill believes.

We can achieve this modification within Mill's framework, because Mill does not argue for an absolute protection of expression. In fact, he notes that considerations of liberty do not categorically override the harm principle. For example, consider the corn dealer case. He argues that speech calling a corn dealer a scammer in front of an angry mob could be restricted, whereas that exact speech would be perfectly protected in a newspaper editorial (52). Thus, he does make temporary exceptions, where he believes that imminent harm would be caused by the speech. Besides the imminent danger case, the harm principle might also apply in cases of libel. The target of the libel person often experiences tangible harm from the falsehood, which means that Mill might argue that libel can be restricted under the harm principle. Thus, given that Mill has

already made these exceptions to the Liberty of Thought and Discussion principle, we can modify his account to include more exceptions without fundamentally changing the framework.

Mill's argument works in an idealized world with unlimited verification resources. In a society where people do have unlimited cognitive capacities, Mill's Liberty of Thought and Discussion would be optimal, because spending resources verifying one statement would not detract from time verifying others. Each statement would have the maximum likelihood of being true, and autonomy would be preserved. However, Mill does not account for the fact that in reality, verifying the truth is costly. The landscape of information and disinformation is difficult to navigate, and realistically, rational humans have limited capacity to evaluate it. In this complex world with fundamentally flawed agents, Mill's ideals of truth and autonomy no longer speak in favor of the same strong free speech protections that he so famously supported.