The Ethics of Political Bots: Should We Allow Them For Personal Use?

JONAS HAEG

University of Oxford

2017 OXFORD UEHIRO PRIZE IN PRACTICAL ETHICS

RUNNER-UP FINALIST: GRADUATE CATEGORY

ABSTRACT

The technology to create and automate large numbers of fake social media users, or "social bots", is becoming increasingly more accessible to private individuals. This paper explores one potential use of the technology, namely the creation of "political bots": social bots aimed at influencing the political opinions of others. Despite initial worries about licensing the use of such bots by private individuals, this paper provides an, albeit limited, argument in favour of this. The argument begins by providing a prima facie case in favour of these political bots and proceeds by attempting to answer a series of potential objections. These objections are based on (1) the dangerous effectiveness of the technology; the (2) corruptive, (3) deceitful and (4) manipulating nature of political bots; (5) the worry that the technology will lead to chaos and be detrimental to trust online; and (6) practical issues involved in ensuring acceptable use of the technology. In all cases I will argue that the objections are overestimated, and that a closer look at the use of political bots helps us realise that using them is simply a new way of speaking up in modern society.

1. INTRODUCTION

A "bot", in the broadest sense, is any computer program built to perform automated tasks. While bots can be embodied in, and control, actual robots, most of them are not and work merely as a software or application on a computer or computing device. The application in charge of making your digital clock update the time on the screen every minute would constitute a bot in this sense, as it merely involves a repetitive, automated task (i.e. changing numbers on the screen). A subset of bots called "internet bots" are automated programs which operate *online* (for instance, gathering information about visitors on websites); and a further subset of these, called "social bots", operate on social media platform such as Twitter. These bots are programmed to run on, or control, their own social media profiles (identical to the ones you and I would have) and their tasks are of the social media type: posting messages, "liking" other posts, "following" individuals, and so on. While some of them are transparently bot-like (e.g. profiles programmed merely to share statistics about each vote in the U.S. Congress), others are designed to *mimic* human behaviour and to be perceived as genuine. Hence, the content of their messages, the patterns of posting, etc. are close to ours: emotional and personal, containing opinionated beliefs, and so on.

Within this latter category of social media bots, we find the bots that are the main focus of the present paper. The term "political bots" refer to human-looking social media bots with a clear *political aim*: to influence political discussions and opinions. For simplicity, I will refer to these bots as "polibots". The use of these bots have increased in recent years. According to some estimates, polibots produced almost 20% of all election-related tweets on Twitter a few weeks prior to the election day of the last U.S. presidential election.[1] Researchers and others warn us that the use of polibots will only increase around the world.[2] As more and more people turn to social media to discuss politics, and use social media as an information source, these platforms are increasingly being dominated by hordes of bots that try to influence opinions. This evolution has attracted academic and political interest from many[3],

1. Bessi and Ferrara (2016).
2. See, e.g., Miller (2017); Guilbeault and Woolley (2016).
3. See, e.g., Oxford's "Computational Propaganda Project": http://comprop.oii.ox.ac.uk/

but few philosophers have engaged with the topic. With this paper, therefore, I hope to start a philosophical discussion about this topic by investigating the ethics of using polibots.

The current academic and popular literature on polibots is highly sceptical of them—and rightly so.[4] Yet this is largely in part due to their isolated focus on only one sort of polibot, or rather, one sort of use of polibots. I distinguish, roughly, between three types of polibot uses. These are the *Malicious Polibots*, the *Personal Polibots*, and the *Saintly Polibots*. Current writings on the topic focus almost exclusively on the first kind. I call them Malicious because they are the sorts of bots devoted to obviously malicious aims such as sharing hate-speech, false information, suppressing certain voices, etc., often with the aim or intention of maliciously steering a certain election. On the other side are the *Saintly* kinds, which are devoted merely to sharing important, true and relevant political information, with no allegiance to particular issues or sides.[5] In one sense, I think the first kind is obviously objectionable and the last kind obviously good. However, this paper is concerned with the ethics of *Personal Polibots*. In short, this refers the use of polibot technology by individual, political agents as an extension of their political voices. This paper provides an ethical examination of such use of the technology. The guiding question throughout the paper, therefore, is whether we should allow private persons to use the polibot technology to enhance their political voices online, or whether even this sort of use of polibots is objectionable.

The first section elucidates further the nature of the polibots we are focusing on here, and provide a *prima facie* case for their permissibility. After that, I present a series of potential arguments against them. Some of these attempt to target polibot use as something inherently problematic. Others focus on the bigger picture and dangerous consequences. In all cases, however, I rebut the arguments with some concessions. The overall conclusion is that the case against Personal Polibots have shortcomings and that polibots should be seen as largely permissible and indeed potentially valuable.

4. See, e.g., Woolley (2016); Newman and O'Gorman (2017), Andrews (2017), Lee (2017), Alfonso (2012), Finley (2015) and Hern (2017) for examples and sceptical voices.

5. Another term for these are "transparency bots".

2. THE NATURE OF POLIBOTS

In the literature, "political bot" is used to refer to a wide range of different technologies and uses of those technologies. Some bots are programmed solely to "repost" and "like" certain other people's posts; and others merely devoted to "follow" certain figures to boost perceived popularity.[6] Others again are programmed to share their own content, and these seem to me to be the ones most clearly programmed to mimic genuine humans. They are relatively easy to set up. One needs to pre-write the content that is to be shared, and then decide how frequently the bot is supposed to post these tweets. Conceivably, they could be programmed to post every few seconds, but most likely people will want them to look more human and therefore lower the frequency of posts and also program in "off-time" to make it look like the bot is sleeping, working, etc.

As said in the introduction, even amongst the bots devoted to share their own content, there are large differences in what sorts of content and aims they can be programmed for. The *Malicious Polibots* are so called because they are obviously problematic. These bots, however, can be argued against on grounds independent of the polibot technology itself. After all, the explicit content and aim of these bots are types of content and aims we find morally objectionable in general. For instance, the sharing of hate-speech, fake news and information, trying to suppress certain voices, and so on. It is easy to see that these are morally problematic independently of the technology. On the other side are the *Saintly Polibots*. And likewise, it is easy to see why we should be in favour of their existence on grounds independent of the technology. After all, they are programmed to share non-biased and politically relevant information, which are often not shared widely in mainstream news sources.

However, my interest here lies in the ethics of *polibots* in themselves. The problems with *Malicious Polibots* and the good features of *Saintly Polibots* aren't problems or features connected to the *polibot* technology *per se*, but rather the specific content or aims they have. To delve deeper into the specific issue concerning the use *polibots* itself, I focus on what I call *Personal Polibots*. These seem to be to be neither obviously objectionable nor obviously morally good. I am assuming that they will not share obviously problematic content such a hate-speech and fake information, and that the aim isn't malicious in the sense of being intended to fool people to sway

6. See, e.g, McKelvey and Dubois (2017) and Wooley and Guilbeault (2017) for more on different kinds of polibots.

their opinion, or try and scare away certain voices online. Rather, the polibots are programmed to share, in a non-offensive manner, information and opinions supporting the political cause of the programmer themself. In a sense, I take them to be programmed extensions of how many people are already using their personal social media accounts to engage in political matters and gain political influence. They are used to share the political beliefs and messages of their programmer. At the same time, this use isn't assumed to be morally saintly either. For in contrast with the objective, transparency-type polibots, normal people aren't neutral. They choose sides, posts information supporting their cause, and might refrain from posting information that would hurt their side. *Personal Polibots*, then, while required not to share hateful and false content, aren't required to be impartial. Focusing on this personal type of polibots will allow us to more clearly investigate the ethics of the technology itself, and whether we should permit its use or not.

Now, given the restrictions on what sorts of polibots we focus on, we can reasonably say that the sort of polibots of interest here are those which function as "communication tools". By this I mean tools that help us enhance our communication in different ways. After all, what is it that the polibots mentioned would add to the ordinary means that political agents have of sharing their political messages? They provide an enhanced scope of reach and potential for influence. Using one personal user profile online in real time gives you a certain level of reach and influence (depending on other factors as well, of course, such as amounts of followers, hashtags, and so on). But with the aid of polibots that can act much quicker and more constantly than a single person can, the scope and influence can be enhanced drastically. As one reaches more persons with one's message, it becomes more likely that one will reach someone who is prone to be influenced by that message as well. Now, this feature of enhancing communication is one that polibots share with many other political "communication tools" available today. Examples include hanging up political posters, writing op-ed pieces for newspapers, paying for political advertisements, and so on. Even using one personal social media account would be an example. As such, the polibots discussed here seems, on the face of it, to simply be a *new* communication tool available to people.

This provides, I think, a *prima facie* case for their permissibility. However, I suspect many people have some reservations also about this sort of use of polibots. Below, I outline and respond to some of the potential objections people might raise.

Although they highlight interesting ethical issues surrounding polibots, they do not, I argue, manage to show that we shouldn't permit them.

3. ARGUMENTS AGAINST THE PERSONAL USE OF POLIBOTS.

3.1. Objection 1: Polibots Are Too Effective.

Someone might argue that polibots are impermissible because they are "too effective" communication tools – that they will have too much influence. One might divide this worry into two separate ones. First, one might worry that there is such a thing as "too much influence" *simpliciter* and that polibots have this. Second, one might instead worry that polibots unfairly give *some* people *more* influence than others have.

The first worry might be put, somewhat metaphorically, like this. Everyone is entitled to a *political voice*, to engage in political matters with and try to convince one another. However, using polibots gives one, effectively, *several* political voices because they enhance one's message so much. But since we are only entitled to *one* voice, their use is impermissible. The first problem with this objection, however, is to explain why polibots go above some sacred threshold of political influence. Presumably, using polibots might be more effective than simply walking around the streets alone telling people your opinion. But there are many other sorts of communication tools that seem much more effective at influencing than this, which we nevertheless do allow. Examples might include writing op-ed pieces for the news, running a blog, or hanging up posters. If these are not above the aforementioned threshold, but still more effective than using no technology at all, it is hard to determine whether polibots are too effective. For showing that it is *more* effective than using no technology doesn't suffice to show that it is *too* effective.

Moreover, the objection is most likely false if we focus on the use of just *one* or a *few* polibots. For the use of only a few polibots will in most cases, I think, not be more influential than the use of many of the other permissible technologies mentioned above. Hence, this worry cannot be an intrinsic objection to polibots. Granted, however, the worry is more realistic when we consider the use of *many* polibots. Even so, showing that some use polibots can be more effective than other traditional communication tools does not entail that this is too effective. For "more than" doesn't entail "too much". Moreover, Woolley and Guilbeault (2017, 10) have already dis-

cussed this potential effect of polibots which they call the "megaphone" effect, i.e. the use of polibots to heavily amplify a political message. They also say that the worry is not really a polibot problem *per se*, but rather a problem with the *number* of polibots. Viewed in this light, however, we see that the worry isn't particular to polibots. For the megaphone worry arises for all sorts of communication tools. Compare paying for TV adverts, phone banking, and so on. All of these tools *can* become objectionably influencing when we start considering large numbers. For instance, we would object to someone buying up all air-time on TV to advertise for his candidate. The problem, however, isn't (necessarily) that we deem political TV adverts themselves impermissible tools, but rather that it is impermissible to exert so much influence. The same can go for polibots: they might not be impermissible in themselves, but their use can become impermissible if too many of them are used. At best, however, this objection would support some sort of regulation on polibots— something we're already familiar with from other communication technologies.

If we move to the second version of the worry, the same response probably hold. Recall, the worry was the polibots can give some people much more political influence than others. But again, this is a worry about the numbers and the scope of the availability of the technology. However, it also bears mentioning that polibot technology might actually have the opposite effect to some extent. People today are often worried that a few wealthy individuals gain disproportionately much political influence. But polibots are extremely easy to set up, and so the technology could potentially help democratize political influence somewhat by making it more available to those who cannot compete with rich persons on other arenas. Of course, though, money might buy one more effective bots than the average person can create themself, and so the unfairness worry might continue as before. Again, though, as I said, this only highlights the potential need for some sort of regulation.

In sum, this objection does not speak forcefully against allowing the personal use of polibots. It points to a general problem with communication technologies, and it is not particular to polibots. Almost any sort of communication technology could conceivably be used so as to be "too effective". This doesn't show that any use of the technology is impermissible, and, moreover, it seems the correct response would be regulation. Lastly, it is worth noting that the *actual* effect and influence polibots have is an empirical question to which we currently lack clear answers. The current research, moreover, has tended to focus on coordinated polibot campaigns and not the

effects of private persons. It might turn out that it will be extremely hard for single individuals to have effects comparable to many of the more traditional means.

3.2. Objection 2: Polibots Corrupt Political Discourse.

Another potential objection is that polibots corrupt political discourse. Political discourse, the argument might go, is by definition limited to political *persons*. By activating polibots, one introduces non-persons into this discourse and therefore corrupts it. This objection is stronger than the above objection because it aims at showing that there is something intrinsically wrong about using even only *one* polibot.

The objection raises an interesting question about the ontology of polibots. In one sense, as I said earlier, a polibot appears as a mere tool or extension of the human controlling them. On the other hand, some might feel that they are more than this. Imagine, for instance, a social media platform of 100 genuine persons engaging in political discourse. Then, one of these people adds a polibot. It now seems as if there are 101 entities engaging on that arena, not all of whom are persons. There is some force to this intuitive picture, but let us attempt to dissect it: what is really the problem with that 101st entity?

Recall, polibots as here defined are very simple. A person needs to write and intend all of its posts beforehand. It is not really an intelligence in itself. Suppose, then, that I had the following two options: (1) stay at home for the weekend and post my regular political tweets, or (2) automate all of this in a script and take the weekend off. Returning to the picture above, it seems that choosing (2) would mean there are 100 entities on the platform, not all of which are persons. Yet, viewed in this light, it seems much less objectionable. I fail to see why the mere *automation* of one's political engagement itself is the problematic feature. The only difference between (1) and (2) seems to be that I am not posting my opinions *in real time*. But intuitively, this lack of real-time engagement seems insufficient to make it impermissible. This form of automatization doesn't, I hold, amount to introducing a distinct, non-person entity. Indeed, it seems that (2) has a lot in common with "auto-emails", i.e. emails typed beforehand and set to send at some later point. But that doesn't strike us as importantly different from writing emails in real time. In short, I cannot see why *automating* one's political engagement in this way corrupts the political discourse.

The opponent might, however, highlight two problems with the picture above. Polibots are different from (2) in two key ways: polibots usually come *in addition*

to one's personal account and have their *own identities*. It is not merely like a script running "your" online life (i.e. your personal profile); it adds seemingly disconnected users to that activity. I grant this. But, if one agrees with the above—that the corruption doesn't come from the mere automatization of political activity itself—it seems as if one must think that the corruption of political discourse comes from these two features: *additionality* and *anonymity*. This is implausible, however. It is already somewhat commonplace for person to have several social media accounts without necessarily disclosing this. Since these actions also involve additionality and anonymity, the current objection would equally apply to these people. However, the objection was that polibot activation are tantamount to introducing non-person entities online. But it seems outright implausible to say this in the case of someone activating users with additionality and anonymity (but which are controlled directly by the person). Clearly, this simply cannot be tantamount to introducing non-person entities; after all, they are themselves controlling these additional, anonymized users in real time.

Could it, however, be that it is only *anonymity* and *additionality + automatization* that results in corruption? Perhaps, though I find it unlikely that the mere, simple automatization could make this much of a metaphysical difference to the ontological status of those user profiles. In sum, therefore, I conclude that one cannot conclude that polibots are corrupting political discourse by introducing non-person entities.

3.3. Objection 3: Polibots Deceive.

Despite my pessimism above, I grant that there is another side to the worry. What matters, some might say, isn't the strict "ontology" of polibots but rather the perception that other people have of them. Using polibots without disclosing them as such, they might say, amounts to a form of *deception*. It is not the explicit content that is deceptive, but rather that people will be fooled by the activity of polibots and come to have false beliefs. In particular, they will believe that the bots are controlled by separate, genuine humans. As such, upon encountering my personal activity and my bot's activity, someone might wrongly infer that she has encountered *two* separate humans online, and in that way is deceived. Lynch (2016, 154-6), for instance, seems to think social bots in general are problematic for these sorts of reasons, claiming that bots are "massive deceit machines" that "operate on getting people to assume they are dealing with someone real who is sincere in their assertions".

However, I believe this worry is overestimated when we are dealing with *Personal*

Polibots. Suppose I have a polibot, B, which tweets the content "Donald Trump has been sued 135 times since taking office. Clearly he's not fit to be president!".[7] Deception is about causing false beliefs, but what belief is relevant here? It might be that *someone believes that Trump isn't fit to be president because of all the lawsuits*, but in a sense this belief is true. After all, *I*, the creator of B, believe so. In fact, it seems hard to spell out exactly what belief we are talking about. But let us grant that there is such a belief. Let us call it the belief that *B is a separate human who believes that* ... so that upon encountering both me and B, a person will make a wrong inference about how many persons believe the mentioned content. Has my activation of B constituted deception?

To deceive is traditionally defined as "to intentionally cause [someone] to have a false belief that is known or believed to be false" (Mahon 2015). On this account, however, it is questionable whether my polibot amounts to deception. For it is not clear that the false belief in question is part of the *intended* effect. It seems to me that the primary intention is not to get others to think that the polibot is a genuine human with certain opinions. Rather, the primary aim and goal is to get another person to access the explicit *information* (in the tweet) and become influenced by it. Of course, causing a (false) belief about who the bot is might be an effective means of doing so. After all, part of an effective way of getting someone to believe that *p* is to get them to believe that s*omeone sincerely believes that p*. But it seems that the agent behind *Personal Polibots* (as elucidated in Section 2) would be satisfied whether or not the other person saw through his bot, as long as she took information seriously and based her opinions partly on it. As such, they would pass the usual test for what is intended. Hence, I think the causing of the false belief will not be what is primarily intended. In many cases, it will be a mere unintended consequence.

Someone might respond that this misses a crucial fact: if people really didn't intend to deceive others about the identities of the polibots, they wouldn't use the polibots described here at all. For those bots are assumed to be anonymous. People who do not want to deceive would either mark their bots as bots or refrain from using the technology altogether. To respond to this point, we should look deeper at the reasons why people who intend to spread their message would nevertheless use these sorts of bots. One potential reason might be this. People want to spread their message as much as possible and so they seek technologies to help them do so. They also know that people are sceptical of bots (for instance, given all the negative press about bots, we are likely to be sceptical about who controls the bots, whether they are speak-

7. See Hallemann (2017).

ing truthfully and so on). They therefore opt for anonymous polibots, because they aren't obviously identifiable as bots (which they would be if they were, for instance, marked as "bot"). These users don't want to use polibots to trick people about how many people support issue *x* and in *that* way convince them to support *x*. They want them to access the information tweeted and become influenced by considering that information seriously. Hence, the main intention and goal behind the act (of activating polibots) seem acceptable, even if effectiveness in reaching the goal (making people access the information and considering it seriously) might require using bots that aren't so easily identifiable as bots.

Of course, the opponent might respond that this nevertheless means that my use of polibots *relies* on tricking people into thinking the bots are real. My plan relies on fooling others even if my ultimate goal seems fine. Hence, it seems that I cannot be completely free of the charge that I am deceiving my peers. In response, however, I think we should draw a relevant distinction. This is the distinction between a plan's relying on *tricking* others into thinking that my bots are genuine humans, and a plan's relying on others not identifying my bots as bots. Take the person above who wants to convince others that *x* is correct by making others believe their pro-*x* bots are humans and in *that* way influence their opinions about x; and compare me, who want people to access the information shared by my bots only and in that way influence their beliefs. Both of our plans, in a sense, relies on people not identifying our bots as bots, but for very different reasons. My plan does so because, given the propensity of humans to be skeptical of bots and their content, my intended goal depends on them not identifying the bots. Call this sort of reliance on false beliefs for *negative* reliance. It is only a reliance on a false belief because the opposite of the false belief will thwart my aim. My plan only relies on anonymity of bots because of people's prejudice against bots. The other person, however, will *positively* rely on false beliefs in the sense that their entire plan is to deceive and in *that* way influence. For them, false beliefs are the means; for me, false beliefs are only contingent, unfortunate requirements. In a sense, the relevant distinction is between contingent and necessary reliance on false beliefs; and the anonymity of the bots serve very different functions in our respective plans—in one case as the means of getting others to support x, in the other case merely to avoid triggering scepticism and counter-productive effects. To me, a plan that only has negative reliance on false beliefs in this way doesn't seem so obviously deceptive.

Moreover, there is always also a separate question, namely about permissibility

all things considered. *Even* if *Personal Polibot* use would constitute some form of deception, we can ask how wrong, *all things considered*, that is. As will become clearer below, there might be countervailing reasons. For the main intention, as I said, is to get people to access information believed to be genuine and relevant to political matters. This is exactly the sort of information and method we *want* people in a democracy to base their opinions on. Hence, if polibots (often) can help boost this method, then the goodness (or democratic value) of this might outweigh the badness of any deceptive side-effects.

3.4. Objection 4: Polibots Impermissibly Influence Opinions.

A worry very similar to the one above is the following. An opponent might say that whether or not we call it deception, the problem with polibot use is this: 1) People are influenced by perceived popularity; 2) using polibots contributes to creating false impression of popularity around candidates and issues; 3) hence, polibots influence political opinions in a deceitful way. Woolley and Guilbeault (2017) have already discussed this sort of phenomena which they call "manufacturing consensus". One can make it seem as if many, many people agree with your beliefs, and in that way influence others' beliefs. Above, I tried to suggest that this won't be the primary intention of those using Personal Polibots. The worry here, however, is that even where such effects are unintended, it is nevertheless morally objectionable if polibots do have these effects. Before considering this objection, we should get clearer on how the problematic effect in question can be produced by polibots.

One direct way is if there are two political candidates, for instance, C and D where D is more popular. If I had control over many polibots and made these share pro-C sentiments, it will become more likely for another person online to encounter a C-supporter than a D-supporter. Over time, this could make others come to believe that C is *actually* more popular than D. In turn, this might influence the opinions that others have of C and potentially make them less confident in D or more confident in C. After all, they might think that all those C-supporters cannot be wrong. Admittedly, however, this effect is questionable. At least research done on the effect that *polls* have on political opinions seems to suggest that few people tend to *switch* sides based on perceived popularity.[8] Even so, since the number of actual supporters in many elections seem to be in the millions it seems unlikely that any single individ-

8. See Snyder (2012).

ual will be able to have this effect on perceived popularity. It seems to be a worry that is more relevant when we consider the aggregate use *Personal Polibots* of many people.

However, it should firstly be noted that this sort of effect would not be particular to polibots at all. Indeed, news agencies might often unconsciously (or consciously, of course) achieve similar effects when they report more on certain candidates over others when this disproportionate attention to certain candidates have no basis in actual popularity. Similarly, if C's supporters are more active and effective in their canvassing efforts, they might get a similar effect on people even if D is actually more popular. Yet canvassing seems to be an activity we find permissible. Secondly, this feared effect is already built into social media and the internet itself because not all citizens are online in the first place. And there is no guarantee that the number of supporters for C and D online will be an accurate representation of the actual numbers either, for D might be more popular amongst more senior voters which tend to be less active online. Hence, basing one's beliefs about candidate popularity on social media is already risky, and the effect (false impression of popularity) is already likely to exist online whether or not polibots exist. Moreover, insofar as the opponent is criticizing the use of polibots for contributing to inaccurate representations of support online, he should also, by the same token, criticize those currently *not* online for doing the same. Yet, this seems like a path we do not want to go down. Third, as having an online presence online becomes more and more common for the entire population, and the polibot technology remains easily accessible, it might be the case that my 100 polibots in favour of C will be counterbalanced by 100+ polibots in favour of D; in which case the effect from polibots would disappear or decrease as the proportions of support would remain constant or at least similar. Lastly, there are also other, and more officially recognized, sources of candidate popularity, such as polls. Although not guaranteed to be accurate, these do make a serious effort to be as accurate as possible, whilst there isn't necessarily any such effort made by social media platforms. This should, hopefully, help balance false perceptions of popularity that might be created online. In sum, then, we can see that the problem of creating false impressions of popularity might not be such a persuasive issue for the bot use that we focus on, nor is it particular to polibots.

There is also, however, a less direct way of influencing opinions which Woolley and Guilbeault (2017) call "agenda setting". One can influence political opinions by controlling what topics and issues are discussed and perceived as important. This also seeps out into the real world, as news stations and others pick up on "hot topics"

online and report on that. Suddenly, it can seem to everyone that some issue is very important, even though it actually isn't. Further, if this issue puts C in a good light in some way, it can make people more inclined to support C as well. In sum, through affecting what is talked about and viewed as important, polibots can change opinions. Although there are admittedly frightening examples of this effect[9], the worry as it relates to *Personal Polibots* seem less serious. Indeed, it seems that many of the responses given above hold with equal force here. To reiterate, as well, it seems unlikely that a single individual (within the confines of Personal Polibots) will be capable of wielding such power. Hence, the agenda setting effected by polibots will largely emerge out of many distinct individuals using their own *Personal Polibots*.[10] It can be hoped that, as more and more would use the technology, one potential effect of this is that the sorts of issues that get set highest on the agenda will mirror the issues that people, in actuality, do put most weight on as well. It is, of course, always a problem that the issues and topics that people will boost with their bot uses will turn out to be fake or deceptive itself—perhaps due to the large amounts of fake news circulating. The problem, here, however is first and foremost a problem about fake news itself, and not necessarily about the polibots. Furthermore, it is also a possibility that the polibots would share and boost genuine news and in that way help countervail the influence of fake news online. Hence, it is very unclear what the effects will be. And it seems to me that, morally, the goodness or badness of this sort of influence on opinions will depend a lot on the source of the information shared and boosted as well.

Even if all the issues boosted were genuine and morally important, the critic might insist that there is a possibility that some will be influenced in the wrong way, i.e. *purely* by the perceived popularity of the case, not the substance itself. It bears mentioning however, that any sharing of information runs the risk of influencing opinions in suboptimal ways. This possibility seems almost impossible to get rid of, with or without, polibots. Moreover, this seems more of a problem with human psychology itself, and not directly the polibot use in question. Even doing our best to reduce this sort of influence, it seems inevitable that someone will be influenced by mere popularity.

Lastly, it is also worth mentioning some potentially positive uses of agenda

9. Woolley and Guilbeault mention "Pizzagate", the conspiracy linking Hillary Clinton to human trafficking and child abuse, which seems to have been boosted and made into a much talked about issue online, partially, by the use of polibots.

10. I am focusing here on aggregates of distinct individual uses. When we start thinking about collectively acting groups and coordinated polibot campaigns, many distinct worries arise.

setting with polibots. Currently, some information about political candidates and certain issues often get less attention in the news, etc. for no "morally" good reason. A damaging report about D's conduct towards women might get little attention because, for instance, D is backed by powerful people who try to keep it out of the discourse, or D is very popular and his supporters do their best to keep the stories out of the online discussions by not engaging with the stories at all. Similarly, a morally and politically important issue might get little attention simply because it is a minority issue. None of these are morally good reasons for their unpopularity. It can be hoped, then, that individuals might be able to use polibot technology to bring such issues and information more to the forefront of discussions, and force people to engage with them more seriously. This seems to me a virtuous use of the technology. Moreover, at least for many realistic such scenarios, it seems as if the desired effect isn't merely to sway political opinions by sheer perceived popularity even though that can be an effective means for, say, an underrepresented group. More centrally, I think, many in such a position would not simply aim to change policies by making politicians believe there is massive support for it, but more directly through getting these politicians (and others) to have a change of heart themselves. In other words, getting people to *realize* that the issue is real and important, and in *that* way get other genuine humans to be morally motivated to support a change. They would seek a genuine recognition of the problem; and that requires changing hearts, not merely pushing through unmotivated change.

In sum, then, I believe the objection is overestimated. I've tried to argue that, insofar as the problem is potentially real, it will not necessarily be widespread on the individual level. Nor is the problem unique to polibots either. Lastly, there also seems to be potentially good uses of the effect in question.

3.5. Objection 5: Collective Polibot Use Creates Chaos and Distrust.[11]

Much of the discussion so far have focus on the individual use of *Personal Polibots*. Let us move instead to more large-scale effects that would potentially emerge if many, or indeed most, started using *Personal Polibots*. There might be a worry that chaos and distrust will increase. Suppose the two candidates C and D have 1 million and

11. This objection is inspired by some brief remarks made by Julian Savulescu during the presentation of a shorter version of this paper in the final round of the 2017 Oxford Uehiro Prize in Practical Ethics competition.

5 million actual supporters online respectively. Suppose that each supporter would use 100 polibots each. Suddenly we go from a pool of 6 million users to, effectively, a pool of 300 million users. That is an extreme increase. It is easy to imagine that chaos would ensue, with the social media platform being dominated by mere polibots talking past each other. As this increases, people will find it harder to find other genuine humans. They will start becoming much more distrustful of users they meet online as they know it is most likely a fake one. As this increases, people will also likely start to look for other less bot-infested platforms and move there, which again will become infested, and people will migrate elsewhere again, and so on. Indeed, polibots might undermine the essential function of social networks: being a virtual space for people to discuss and engage with each other.

Despite the dystopian feel to this scenario, I believe it would most likely be averted in a natural way. For instance, just like websites limit adverts and email services limit spam emails, it could be imagined that companies like Twitter and others would regulate the total number of polibots permitted on their sites at any one time, or would attempt to regulate in some non-biased way the number of polibots that are *shown* in people's feeds. It is also likely that user themselves might act so as to decrease the "chaos". We are all already familiar with "blocking" technology, in relation to adverts and spam emails. It seems to me likely that "bot-blocking" technology that decrease the number of bots visible in one's feed would emerge as the use of Personal Polibots increase as well. Moreover, it might be that such a natural response is to be preferred over a desire for a complete eradication of *Personal Polibots*.

The reason is this. Most blocking technologies are seldom 100% effective, as people find different ways to trick them. It is likely therefore that some polibots will get past a person's blockers in any case. This is important as most people online do find themselves in self-imposed echo chambers and informational or filter bubbles. [12] These embed us in social media feeds containing information the algorithms believe we already approve of. This is a democratically unfortunate effect as people are less likely to be confronted with opposing information, arguments and views. If the polibots outside these bubbles are many enough, however, it is more likely that some will seep through the both the filter bubble and the blockers, and thereby help confront people inside them with relevant information that wouldn't otherwise easily come across. After all, even inside a blocking-enforced bubble it is hard to hide from very popular and widely spread issues. And the more users appear to discuss a certain

12. See Hossain (2016)

issue, the more likely it is that one of these users—whether genuine or a bot—will slide into one's feed, past the bubble and the blockers.

In sum, then, I believe there might be responses that emerge naturally to combat any feared chaos and distrust online. These natural responses might moreover be favourable to a complete eradication of the bots themselves. The more people are talking about a certain issue and using polibots to boost that discussion, the more likely it is that others will be confronted with those issues (without encountering the chaos), even hiding behind bubbles and blockers. Indeed, the result might be that people are more exposed to the information that matters the most.

3.6. Objection 6: Permissible Use of Polibots is Practically Impossible.

As we've seen throughout the paper, I have only defended a very limited type of use of polibots. Plausibly, we might require that bots do not share hate-speech, do not spread deceitful information, isn't used merely to suppress others, and so on. Potentially, we might also need some form of regulation to avoid that some people get disproportionality many polibots. However, it seems very hard to ensure that these requirements will be obeyed. To ensure this, we need technology that is capable of identifying both bots (and who controls them) and their content. Yet, this will be extremely difficult to achieve. Hence, fearing that "bad bots" will thrive unless we have this technology, it is overall more advisable simply to discourage the use of polibots, and indeed work to shut all of them down. This objection, then, accepts the claims of this paper but argues that there is no way of practically allowing these sorts of bots without having an increase in bad forms of bot use as well.

The problem with this argument is that it is overly pessimistic about technology. For instance, the fact that researches are able to analyse the numbers and activities of polibots, and indeed their origin, shows that we already do have some methods available to detect bots. Moreover, we already know that companies like Google and Facebook are already developing automated technologies that can detect fake news, hateful comments, and so on.[13] It doesn't seem to me, therefore, that technology capable of ensuring compliance with the requirements on polibots like the ones mentioned above is too hard to engineer.

13. See, e.g., Feldman (2017), Leong (2017), and Frier (2017).

4. CONCLUSION

This paper has attempted to highlight and discuss some initially intuitive worries regarding one potential use of polibots, namely the personal use of polibots to enhance one's political voice. Although this discussion has helped clarify some ethical issues, all of the putative objections suffer from weaknesses. Many of them are very general and would equally apply to technologies we do find permissible; some of them clearly fail; and some of them are very likely overestimated in their force. Moreover, I've also tried to highlight some potentially positive and valuable effects of polibot use.

In sum, I've tried to outline a plausibly permissible use of polibot technology which I think escapes much of the intuitive worries we might have about them. At the end of the day, I fear that most of the scepticism towards this use might arise because of the word "bot" itself. As we look deeper down at the potential problem areas, however, we see more clearly to what extent this use of the technology is just a new way of speaking up in today's world—the kind of technology we've sought since the beginning of democracies, and even before that.

However, we should keep in mind that although I have played the devil's advocate on behalf of *Personal Polibot* use as a mere extension of one's political voice and agency, there are many other areas concerning different polibot uses (such as those by governments and official campaigns), related activities (such as the commercial selling of polibots), future polibot technology (such as more autonomous ones) that deserve their own ethical examinations.

REFERENCES

Alfonso, F. (2015). *Twitter Bots Silence Critics of Mexico's Leading Presidential Candidate*, from The Daily Dot: https://www.dailydot.com/news/pena-nieto-twitter-bots-mexico-election/ [Accessed 25 July 2017].

Andrews, C. (2017). *Are Political Social Media Campaigns a Threat to Democratic Elections?*, from E&T: https://eandt.theiet.org/content/articles/2017/05/are-political-social-media-campaigns-a-threat-to-democratic-elections/ [Accessed 17 May 2017].

Bessi, A., & Ferrara, E. (2016). Social Bots Distort the 2016 U.S. Presidential Election Online Discussion. *First Monday*, 11(21).

Cook, J. (2011). *Update: Only 92% of Newt Gingrich's Twitter Followers are Fake*, from Gawker: http://gawker.com/5826960/update-only-92-of-newt-gingrichs-twitter-followers-are-fake [Accessed 25 July 2017].

Feldman, B. (2017). *Can Google Use AI to Fix the Comment Sections?*, from NY Mag: http://nymag.com/selectall/2017/02/google-introduces-perspective-a-tool-for-toxic-comments.html [Accessed 13 August 2017].

Finley, K. (2015). *Pro-Government Twitter Bots Try to Hush Mexican Activists*, from Wired: https://www.wired.com/2015/08/pro-government-twitter-bots-try-hush-mexican-activists/ [Accessed 25 July 2017]

Frier, S. (2017). *Facebook Automates Effort to Flag 'Fake News' for Fact Checking,* from Bloomberg: https://www.bloomberg.com/news/articles/2017-08-03/facebook-automates-effort-to-flag-fake-news-for-fact-checking [Accessed 13 August 2017]

Guilbeault, D., & Woolley, S. (2016). *How Twitter Bots are Shaping the Election*, from The Atlantic: https://www.theatlantic.com/technology/archive/2016/11/election-bots/506072/ [Accessed 25 July 2017].

Hallemann, C. (2017). *Here's What You Need to Know About Donald Trump's Lawsuits*, from Town and Country Magazine: http://www.townandcountrymag.com/society/politics/a9962852/lawsuits-against-donald-trump/ [Accessed 13 August 2017]

Harn, A. (2017). *Facebook and Twitter are Being Used to Manipulate Public Opinion*, from The Guardian: https://www.theguardian.com/technology/2017/jun/19/social-media-proganda-manipulating-public-opinion-bots-accounts-facebook-twitter [Accessed 25 July 2017].

Hossain, I. (2016) *Filter Bubbles are Shrinking Our Minds*, from Huffington Post: http://www.huffingtonpost.in/ishtiaque-hossain/filter-bubbles-are-shrinking-our-minds_a_21469747/ [Accessed 27 August 2017].

Lee, T. Y. (2017). *Bots Used to Bias Online Political Chats,* from BBC News: http://www.bbc.com/news/technology-40344208 [Accessed 25 July 2017].

Leong, L. (2017). *Fighting Fake News: How Google, Facebook and Others are Trying to Stop It*, from Tech Radar: http://www.techradar.com/news/fighting-fake-news-how-google-facebook-and-more-are-working-to-stop-it [Accessed 13 August 2017].

Lynch, M. (2016). The Internet of Us: Knowing More and Understanding Less in the Age of Big Data. New York: Liveright.

Mahon, J. (2008). *The Definition of Lying and Deception*, from Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/entries/lying-definition/ [Accessed 25 July 2017].

McKelvey, F., & Dubois, E. (2017). Computantional Propaganda in Canada: The Use of Political Bots. Working Paper 6, Oxford University, Project on Computantional Propaganda.

Miller, C. (2017). *Governments Don't Set the Political Agenda Anymore, Bots Do*, from Wired: http://www.wired.co.uk/article/politics-governments-bots-twitter [Accessed 25 July 2017].

Newman, H., & O'Gorman, K. (2017). *Political Bots are Poisoning Democracy - So, Off With Their Heads*, from The Conversation: https://theconversation.com/political-bots-are-poisoning-democracy-so-off-with-their-heads-79779 [Accessed 25 July 2017].

Snyder, B. (2012) *How polls influence behavior*, from Stanford Graduate School of Business: https://www.gsb.stanford.edu/insights/how-polls-influence-behavior [Accessed 24 january 2017].

Woolley, S. (2016). Automating Power: Social Bot Interference in Global Politics. *First Monday*, 4(21).

Woolley, S., & Guilbeault, D. (2017). *Computantional Propaganda in the United States of America: Manufacturing Consensus Online*. Working Paper 5, Oxford University, Project on Computational Propaganda.