Social Media, Emergent Manipulation, and Political Legitimacy

Adam Pham [ORCiD: 0000-0002-9565-4562]

Alan Rubel [ORCiD: 0000-0002-7028-7305]

Clinton Castro [ORCiD: 0000-0003-4740-0055]

Abstract:

Psychometrics firms such as Cambridge Analytica (CA) and troll factories such as the Internet Research Agency (IRA) have had a significant effect on democratic politics, through narrow targeting of political advertising (CA) and concerted disinformation campaigns on social media (IRA). It is natural to think that such activities manipulate individuals and, hence, are wrong. Yet, as some recent cases illustrate, the moral concerns with these activities cannot be reduced simply to the effects they have on individuals. Rather, we will argue, the wrongness of these activities relates to the threats they present to the legitimacy of political orders. This occurs primarily through a mechanism we call "emergent manipulation," rather than through the sort of manipulation that involves specific individuals.

Psychometrics firms such as Cambridge Analytica[1] (CA) and troll factories such as the Internet Research Agency (IRA) have had a significant effect on democratic politics, through narrow targeting of political advertising (CA) and concerted disinformation campaigns on social media (IRA).[2] It is natural to think that such activities manipulate individuals and, hence, are wrong. Yet, as some recent cases illustrate, the moral concerns with these activities cannot be reduced simply to the effects they have on individuals. Rather, we will argue, the wrongness of these activities relates to the threats they present to the legitimacy of political orders. This occurs primarily through a mechanism we call "emergent manipulation," rather than through the sort of manipulation that involves specific individuals.

We begin by examining two cases. The first is the 2010 Cambridge Analytica "Do So!" campaign, which aimed to tip the balance of a closely contested election by promoting youth apathy in the (ethnically split) Trinidad and Tobago elections. The second is a suite of campaigns by the IRA, which involved the organization franchising its activities to evade detection.[3] Next, we develop and discuss the concept of emergent manipulation, explaining how it differs from other scholarly accounts. Then, we argue that the presence of this sort of manipulation in electoral politics threatens the legitimacy of the elections themselves. Legitimacy, we argue, requires that a citizenry be unmanipulated in a holistic way, independently of whether individuals are manipulated and have their autonomy undermined

1. Manipulation campaigns around the world

1.1. Cambridge Analytica and the Do So! campaign

Cambridge Analytica has become infamous for its involvement in the 2016 US elections and the Brexit referendum, but more recent reports have revealed that the reach of the political consulting and marketing firm has extended far beyond the US and the UK. Alexander Nix, the former CEO of CA, was caught undercover bragging about extortion operations in Sri Lanka.[4] Though Nix's penchant for exaggeration is well-known,[5] a brochure obtained by the BBC revealed schemes in Nigeria, Latvia, and several Caribbean nations, among them Trinidad and Tobago.[6]

Like many post-colonial societies, Trinidad and Tobago has faced deep ethnic divides since the departure of its colonial government. Although inter-ethnic relations are cordial in public, cultural differences and weak institutions have led to professional segregation and a clientelist political system. The primary divide is between Indo-Caribbeans—who tend to support the United National Congress party—and Afro-Caribbeans—who tend instead to support the People's National Movement party—with neither ethnic group owning a majority allowing it to claim durable political control.[7] In such a political climate, where elections are always bound to be closely contested, the sort of manipulative practices associated with CA can be not just influential, but decisive.

Trinidad and Tobago's 2010 elections, which were highlighted in detail by the 2019 Netflix documentary *The Great Hack*, provide an illuminating case study of CA's techniques.[8] The crux of CA's intervention into the elections involved capitalizing on an opposition movement called "Do So." The movement began when a disaffected pensioner, Percy Villafana, refused to allow the then-prime minister to traverse his property during a political walkabout, with Villafana's arms crossed in defiance of the stunt. The movement, which came to be branded by an emblem of crossed arms, went viral on Facebook and soon attracted the attention of CA, which began to bolster the movement via astroturfing efforts in the form of an "ambitious campaign of political graffiti" that "ostensibly came from the youth."[9]

CA's own promotional web materials painted their influence in that election as decisive, arguing that "the employment of CA's research-based differential campaigns and establishment of consistent policy and variegated communications contributed to the [United National Congress[10]￼ Their strategy, more plainly, involved increasing political apathy among *all* young people in Trinidad and Tobago, while anticipating that this would differentially depress voter turnout among Afro-Caribbean youth, relative to their Indo-Caribbean peers. In audio from a sales presentation, Nix himself is strikingly candid about the strategy:

> There are two main political parties, one for the blacks and one for the Indians. And, you know, they screw each other. So, we were working for the Indians. We went to the client and we said, 'We want to target the youth.' And we try and increase apathy. The campaign had to be non-political because the kids don't care about politics. It had to be reactive because they're lazy. So, we came up with this campaign which was all about: Be part of the gang. Do something cool. Be part of a movement. And it was called the 'Do So!' campaign. It means 'I'm not going to vote.' 'Do so! Don't vote.' It's a sign of resistance against, not the government, [but] against politics and voting. … We knew that when it came to voting, all the Afro-Caribbean kids wouldn't vote, because they Do So! But all the Indian kids would do what their parents told them to do, which is go out and vote. They had a lot of fun doing this, but they're not going to go against their parents' will. … And the difference in the 18- to 35-year-old turnout was like 40%. And that swung the election by about 6%, which was all we needed in an election that was very close.[11]

Following the release of *The Great Hack*, officials in the People's National Movement called the legitimacy of the election into question.[12]

Whatever threat to legitimacy the campaign might have caused, the threat did not appear to operate through direct affronts to anyone's autonomy or quality of agency. This, in turn, provides the grounds for deniability. To this end, Nix offered a statement in response to allegations of election manipulation, in which he claimed that "[t]he objective of this campaign was to highlight and protest against political corruption," that "[t]here is nothing unlawful or illegal about assisting with this activity," and that "[CA] has never undertaken voter suppression and there is no evidence to the contrary."[13] Taken at face value, his argument is surprisingly difficult to resist. Since the Do So! campaign *did* begin in a grassroots fashion and *was* furthermore supported by a broad coalition of youth voters, CA's activities cannot be viewed as involving the outright fabrication of a social movement. Rather, we must view these activities as a distorted amplification of an existing movement.

Regardless of its relation to other movements, the Do So! case has several interesting features. First, no individual person or persons were targeted for behavior modification by CA; second, no one's autonomy was necessarily undermined (though someone's might have been); third, there was no publicly disclosed source of central influence. What matters here is that the kind of manipulation we are addressing need not turn on any individual being affected enough for them to lose autonomy.

1.2. The Internet Research Agency and "active measures"

Cambridge Analytica is best known for its connections to the 2016 U.S. presidential election and the UK Brexit referendum. Christopher Wylie (and similar accounts) describes various interactions between Cambridge Analytica and Russia (e.g., testing social media messaging about Vladimir Putin, campaigns for Lukoil, and relationships with pro-Russia factions in the Russia-Ukraine conflict). As a result, Cambridge Analytica's actions in 2016 are often conflated with direct Russian involvement. CA denies any connection to Russian state actors, admitting only to working for private interests in Russia.

The most notable example of Russian "active measures" in U.S. presidential politics is by the Internet Research Agency. The IRA is a Russian state-supported influence operation, described by DiResta et al. as a "sophisticated marketing agency." It has trained and employed "over a thousand people to engage in round-the-clock influence operations" to influence citizens, social organizations, and political processes in a range of countries, including Russia, Ukraine, and the United States. In February 2018, the U.S. Department of Justice indicted the IRA and several of its principals (all of whom are Russian nationals) based on the results of the investigation into Russian interference in the 2016 election conducted by Special Counsel Robert Mueller. The charges in the indictment include conspiracy to commit fraud, wire fraud, and bank fraud. [14]

The activities underlying the charges are social media disinformation campaigns, known as "active measures." The IRA and its agents engaged in a years-long operation to understand U.S. politics and its points of conflict (including agent visits to the United States under false pretenses in order to better understand political culture). They created interwoven networks of ersatz social media profiles and groups that appeared to have a large and "organic," unplanned presence. The IRA purchased ads on social media sites that were targeted at users likely to follow the fake profiles and join the fake groups.[15] The IRA then used these networks to seed and promote inflammatory, divisive content. Notably, the IRA did not focus on any particular ideology or political affiliation. Rather, it sought to engage and enrage social media users from a broad swath of U.S. political positions. The IRA did focus particular attention on Black Americans, targeting this group with ads, creating groups that appeared affiliated with racial and social justice, targeting ads toward places with large African American populations, and focusing on issues that divide Americans along racial and ethnic lines. The Senate Select Committee on Intelligence writes that "[b]y far, race and related issues were the preferred target of the information warfare campaign designed to divide the country in 2016."[16] So, for example, the IRA created social media pages and groups such as "Blacktivist" and posted to social media comments about Colin Kaepernick and other athletes' kneeling protests and about police shootings of Black people.[17]

The pattern of finding groups receptive to provocative, negative rhetoric extended across a broad range of social, cultural, and political affiliations. Some efforts appealed to nativism ("Stop All Immigrants," "Secured Borders"), others targeted messages toward racial and ethnic minorities ("Black Matters," "United Muslims of America"), and some aimed to exploit other cultural and political divides ("Tea Party News," "Don't Shoot Us," "LGBT United"). [18] It is difficult to determine the magnitude of effects these efforts had. However, U.S. Department of Justice reports that the IRA's accounts "reached tens of millions of U.S. persons" and had "hundreds of thousands of followers."[19]

Moreover, the IRA's social media accounts' effects went beyond online viewing. They were the basis for organizing rallies in-person, for recruiting political activists to engage in organizing, and for promoting content promulgated by the IRA.[20] In a study of social media and misinformation, members of the Oxford Internet Institute found that in 2016, prior to the U.S. presidential election, "Twitter users got more misinformation, polarizing, and conspiratorial content than professionally produced news."[21]

The IRA's activities during the 2016 election cycle ranged across social media platforms, including Twitter, Facebook, Instagram, and YouTube. It did not limit its targets to particular political or social orientations, but instead aimed to influence a broad range of views. And while it's precise aims remain unclear, its tactics include influencing users to refrain from voting, to support and vote for third parties, to diminish overall voter participation, to undermine support of political leaders generally, and to build support for "Brexit-style" movements for states (e.g., Texas and California) to secede.[22] Similarly, the IRA sowed distrust in traditional news media by seeding Russian disinformation stories in news media. [23]

There is no definitive information connecting the IRA to Cambridge Analytica, but that is not crucial for our argument here. What matters for our purposes is that several things occur in close, mutually-reinforcing order. First is massive data collection based on lack of privacy protections in social media environments (and in particular on Facebook), the increasing power of data analytics that can use the data collected to better target influence campaigns, and automated systems that recommend how clients can target advertising and which promote content to social media users. The precise relationship between Cambridge Analytica and the IRA may be important for determining responsibility or legal liability, but it is not key in understanding manipulation in the sense we are addressing here.

In addition to the connection between CA and the IRA being unclear, the efficacy of their efforts (individually or collectively) is unclear. Election and policy outcomes are complex phenomena and it is impracticable to identify a single set of events as their cause. And even so, it is unclear whether tactics like those of CA and the IRA are effective at all. According to Kogan, media accounts exaggerate the effectiveness of data analytics and social media campaigns generally, and in particular "[w]hat Cambridge has tried to sell is magic."[24] During the 2016 Republican party primary, the Ted Cruz campaign maintained that its data-driven tactics drove its victory in the Iowa caucus.[25] That view changed as the primary campaign unfolded, with the Cruz campaign growing skeptical and eliminating its use of psychological profiling after it lost the South Carolina primary.[26]

Yet, there is a growing body of evidence for the effectiveness of psychological targeting. ~~[27]~~In particular, a team of psychologists has recently argued that the CA case “illustrates clearly how psychological mass persuasion could be abused to manipulate people to behave in ways that are neither in their best interest nor in the best interest of society.”[28] At the same time, Nix’s cynical argument looms large: there is nothing unlawful, illegal, morally objectionable, or necessarily even manipulative about directing people’s attention to information about corruption. To understand how and why such activities could threaten the political legitimacy of otherwise legitimate governments, we must first understand how the activities are manipulative.

The actions of Cambridge Analytica and the IRA surrounding Do So!, Brexit, and the 2016 U.S. presidential election are in some sense old news. The 2020 presidential election has seen more homegrown misinformation campaigns. Among the most successful of these has been the false claims that states had voting irregularities. These claims have been extensively litigated, and the political pressure for election officials to throw out vote tallies were ultimately unsuccessful. However, a surprisingly large portion of the population took the claims seriously. And this campaign led directly to a violent assault on the U.S. Capitol building that sought to prevent the U.S. Congress from accepting the electoral votes from the states. Indeed, the misinformation campaign has convinced many Americans that the election was illegitimate, and is underwriting a number of actions to restrict voting access in many U.S. states. The 2020-21 campaigns are still unfolding, and analyzing them in depth now is premature. However, we can note here that the same kinds of emergent processes we discuss in this paper are present in 2020-21.

2. The forms of manipulation

2.1. Disputes about manipulation

The philosophical literature on manipulation is rife with scholarly debates about its nature, its extent, and what, if anything, makes it wrong. Is manipulation an effect, an act, or an event? Is manipulation constitutively wrong—applying only to morally unjustifiable conduct—or is it merely usually wrong? How can manipulation be distinguished from similar, possibly overlapping practices, such as coercion and persuasion? Which specific activities—online or offline—count as manipulative? Finally, precisely what values are undermined by manipulative conduct? These are important debates, but we are not going to take a determinate position on most of them. A range of conceptions of manipulation are compatible with the arguments we make below. Whether we conceive of it as overlapping with coercion, or whether we demarcate it from coercion in terms of a distinctive sort of harm, trickery, or carelessness that sets it apart from coercion, the downstream implications of emergent manipulation on issues of legitimacy remain largely the same.

The literature on manipulation most often links its wrongness (if and when it is wrong) to impingements on autonomy, which we will here understand in terms of a capacity for self-government.[29] One way to understand the relationship between manipulation and legitimacy is grounded in the close link between manipulation, the loss of individuals' autonomy, and the implications of this loss on the possibility of democracy. Such an argument works in the following way: If the citizens of a community face a sufficiently strong affront to their capacities for autonomy, they will be left unable to live up to an important civic responsibility, which involves being an informed, conscientious citizen genuinely capable of holding the government democratically accountable. Each of them must be able to critically assess the government's activity and then mobilize accordingly—either in support of the good or in rejection of the bad—or the community will lack a crucial mechanism of democratic accountability. No government can act efficiently unless its citizens can carry out this responsibility, rising to the challenge of holding a government responsible. So, the effects of the IRA and CA's activities at scale is a weakened civil society, rendering effective and responsible government more difficult to achieve (if not impossible altogether). In short, since carrying out one's responsibilities to support civil society requires exercising one's capacity for autonomy, diminishing people's autonomy undermines their ability to underwrite democratic legitimacy to laws, policies, and government actions. Manipulation of this sort makes legitimacy impossible.

Yet, strictly speaking, this argument does not neatly apply to most cases of interest. Not all manipulation has the effect of undermining autonomy, or is even harmful. Consider, for instance, apps such as StayFocusd, which allow users to restrict or control their own access to sites and platforms.[30] To be sure, examples of extreme destruction of autonomy can be found (and appear to be gaining prominence in some online communities),[31] but this model is, in our view, incomplete. Most election-oriented manipulation is not best understood as deeply affecting the autonomy of any one individual social media user, and the degree to which the IRA and CA campaigns affected any one individual person's autonomy was almost always low. 126 million people—the number exposed to IRA-backed content on Facebook—were not epistemologically incapacitated simply in virtue of having seen IRA-backed content. Even if some of the disaffected youth voters in the Do So! case were simply manipulated, this would not explain the drag on legitimacy posed by the Do So! Campaign. This is because the manipulation involved was independent of whether some youth voters were individually manipulated or had their individual autonomy undermined.

Several authors in this volume discuss aggravating factors which appear to make online manipulation more pernicious than manipulation in its more traditional, offline form. It is finely targeted, it exploits dark patterns, and so on. In this chapter, we add another: the practices we discussed in the previous section are examples of what we call "emergent manipulation," which occurs (and matters morally) primarily at the population level.

Here, we adapt the "careless influence" account of individual-level manipulation from Michael Klenk to provide an account of group-level manipulation.[32] Specifically, a manipulator (*M*) aims to manipulate a group (*G*) when:

(a) *M* aims for *G* to perform some act (φ) through the use of some tactic (t), and
(b) *M* disregards whether t reveals eventually existing reasons for *G* to φ.

Klenk's focus is on the manipulation of individuals, and he claims that a key feature of manipulation is carelessness: manipulators are not appropriately sensitive to the reasons of those they manipulate. Our focus is different in two ways. First, we are interested in group-level manipulation. Second, and more importantly, we are interested in a particular *type* of group-level manipulation, viz., emergent manipulation, which involves three additional features. One is that it is *holistic*: it cannot be reduced to the manipulation of individuals. A second is that it is *multiply realizable*: it does not depend on the identities of any specific individuals within the group, but can be instantiated by many distinct combinations of those individuals. And third, it involves *distinctive group-level powers and regularities* which do not appear at the level of the individual, such as the mobilization of a social group.

Next, we will distinguish two types of emergent manipulation, and we will discuss each in turn.

2.2. Stochastic manipulation

One type of emergent manipulation, we will call “stochastic manipulation.” This involves interventions in which no individual is specifically targeted for intervention, and no individual is (or few individuals are) affected so much that their autonomy is undermined. Such practices do, of course, affect some individuals, but they do not affect (or intend to affect) any individual very much, because the intended effect is at the population level. As we see it, stochastic manipulation has two essential features:

1. The approach to the intervention is *dragnet*; it makes initial contact with *many* people but is predicated on the assumption that the behavior of only a *few* will be modified.
2. The aim of the intervention is *marginal*; only relatively few people’s behavior needs to be modified to obtain the desired effect.

In addition to these essential features is an additional feature that bolsters the effectiveness of the intervention:

3. The content of the intervention is *seductive*; those who receive it might already be inclined to agree with it.

2.3. Fragmented manipulation

Another form of emergent manipulation, we will call “fragmented manipulation.” This involves interventions in which there is no openly centralized source of influence, and the manipulation is distributed through more localized (and perhaps unwitting) third parties, such as social media influencers. The features of fragmented manipulation are:

1. The approach to the intervention is *distributed*; those who receive it do not receive it from its actual originator, but receive it through a more localized trusted source.
2. The appearance of the intervention is *misleading*; the intervention appears to be associated with a genuine social movement but has in fact been produced by a centralized group with a disguised agenda, redirecting support from the genuine movement to an ersatz movement.

Though the two forms of emergent manipulation are different (and they can occur at the same time), what makes them morally significant in this context is their intended effect, which is to increase mistrust. Those who receive emergently manipulative interventions are nudged to lose trust either in their fellow citizens or in prevailing institutions. As we will see next, the effect that these sorts of interventions have on social trust can, under the right conditions, play a delegitimizing effect on governments themselves.

3. Emergent manipulation and drags on legitimacy

In this section, we address some of the moral considerations surrounding emergent manipulation. We argue that the phenomenon can, in some cases, serve as a drag on the legitimacy of a political order (regardless of whether that order would otherwise be legitimate).

Following Fabienne Peter,[33] we see two possible sources of legitimacy for political authorities. One possible source of legitimacy flows from the assent of the democratic will, meaning, as Peter puts it, "how well [the authority] can adjudicate between the potentially conflicting wills of the citizens." We will call this sort of legitimacy "democratic legitimacy." Some theorists describe this criterion of legitimacy in terms of public reason,[34] while others describe it in terms of civic participation,[35] but in general, this sort of legitimacy is premised on Rawls's idea of citizens as "self-originating sources of valid claims,"[36] whose claims carry moral weight simply in virtue of having been issued from an autonomous will.

A second possible source of legitimacy, Peter argues, involves a higher sort of normative authority to make binding decisions. On this "epistemic" understanding of legitimacy, legitimate policies are those that are "appropriately responsive" to justified beliefs about what should be done.[37] Joseph Raz's "service conception" of authority exemplifies this epistemic source of legitimacy: on this view, duties to comply with authorities can arise when a subject "is likely better to comply with reasons which apply to him" by "accept[ing] the directives of the alleged authority as authoritatively binding and tr[ying] to follow them, rather than by trying to follow the reasons which apply to him directly."[38] The exercise of authority over someone, in other words, is justified when that authority is exercised in service of the reasons that person already has. This second way of understanding legitimacy allows some space between what is dictated by the democratic will and what can be regarded as politically legitimate.

In this section, we will argue that emergent forms of manipulation drag on both democratic and epistemic sources of legitimacy.

3.1. Affronts to democratic legitimacy

There is considerable disagreement among scholars of democracy, both about what genuine democracy *is* and about what the value of achieving it might be. We might formulate democracy in direct terms—that is, in terms of majority rule or unanimous consent—or indirectly—in terms of satisfying certain deliberative mechanisms. And we might regard the value of democratic decision-making as instrumental—that is, democracy is useful insofar as it facilitates good outcomes—or we might think that certain procedural features of democratic politics inherently confer legitimacy on the decisions it produces. In any case, democratic politics always has the same basic aim: to adjudicate the conflicting wills of the citizens in service of promoting the common good. Achieving this aim is the key to democratic legitimacy. The challenge, then, is that—contrary to Rawls—it is not plausible to think that people are, in general, self-originating sources of valid claims. Rather, people are often manifestly ignorant, irrational, or unreasonable, and it is difficult to avoid the conclusion that this ignorance, irrationality, and unreasonableness can extend into the political domain.

There is more than one way of viewing the source of democratic legitimacy. One way, often associated with Rousseau, involves the idea of a holistic “general will.” According to this sort of view, the common good—which is revealed by but not constituted by deliberative processes—is taken to be distinct from the interests of any individual citizen. Another way of viewing the legitimacy-conferring character of democracy focuses on the structure of the deliberation itself. Josh Cohen, for instance, regards a deliberative procedure as offering legitimacy when the procedure satisfies certain conditions: when it constitutes an ongoing and independent association with final authority, characterized by mutual respect, transparency, and value pluralism, with no suggestion that the results of this process somehow lie apart from the wills of individual citizens.[39]

Regardless of whether we view the citizenry holistically or as merely aggregative, successfully executing the deliberative processes of the sort outlined by Cohen still requires a citizenry that has achieved a kind of collective autonomy that stands apart from the interests, preferences, desires, or values of any one citizen. Several of Cohen’s conditions refer not to the capacities of any one individual within the democracy, but to an irreducibly population-level property: its degree of social trust. The way to understand this property, in turn, is in terms of collective autonomy.

Scholars, of course, have long disagreed about the nature of *individual* autonomy. Some, such as John Christman, understand autonomy as, at bottom, a matter of how individuals’ internal motives relate to their history and psychology, while others, such as Marina Oshana, understand autonomy as fundamentally relational.[40] Setting aside issues related to collective competence and collective relations for a moment, we can see that the crux of collective autonomy involves what we might think of as “collective authenticity.” This is the extent to which a collective would not be, in Christman’s terms, “alienated” from a given decision “upon (historically sensitive, adequate) self-reflection.” To satisfy this nonalienation condition is to feel and judge that the decision could “be sustained as part of an acceptable autobiographical narrative.”[41]

Groups, or collectives, can be alienated from their decision-making just as individuals can. To illustrate this notion of collective alienation, we might imagine an assembly of individually well-informed, rational, and reasonable citizens, who all share an agenda of supporting some sort of public good, such as the construction of a public school, park, or healthcare clinic. However, suppose that the collective lacks adequate social trust, at least in the sense that vague rumors abound throughout the community about “free-riders,” leading each of the assembly members to reasonably question the motives of the others, and thus, to question the ultimate practicality of the agenda itself. The failure here involves a lack of common knowledge within the collective, rather than a shortcoming on the part of any individual. This is because although everyone can (by hypothesis) be counted on to contribute to the good (or at least to behave according to some norm of reciprocity) even in the absence of external enforcement, none of the citizens are in a position to reasonably can reasonably believe that they can count on their fellow citizens in this way. Whatever its merits might be, the policy lacks democratic legitimacy.

There is more than one way to interpret this collective failure. We might interpret it robustly; in terms of, say, a failure to form the “joint intention” implied by each of their individual views.[42] Or we might maintain a more individualist outlook, arguing that the assembly doesn’t “really have any moral status” but that the “distinctively collective interests of individuals mean we must, in some respects, act as if” it does.[43] The key point is that on either interpretation of the failure, the moral of the story is stark: since it is (individually) rational for each member to contribute nothing to the (presumed to be hopeless) public good, everyone voting their own individual interests is a highly stable equilibrium, meaning that no single assembly member would have an incentive to change their voting. It is difficult to imagine a collective that is more alienated: the assembly will not be able to support its own stated agenda despite the unanimous support of that policy from its members.

For an assembly in a complex democratic society to function appropriately—or even get off the ground—it must holistically embody some degree of mutual trust. Within a group, a collective lack of trust functions as a drag on the democratic legitimacy of any group proposal they might consider together: it would be reasonable for any of the assembly members to vote down the proposal. As we will see in section 3.3, one of the primary effects of the CA and IRA campaigns is to undermine the basis of that trust without violating anyone's individual autonomy.

3.2. Affronts to epistemic legitimacy

At first, it might not be evident that there could ever be any source of normative authority apart from that which flows from the will of the people (at least indirectly). How, in a genuine democracy, could there ever be "sufficiently justified beliefs about what should be done" that depart substantively from what the governed themselves have consented to? What kinds of parties could have the standing to interfere with a genuinely democratic decision? And what kinds of issues could be at stake in such cases?

Peter, for instance, offers "[p]olitical decisions that sanction unnecessary harms to small children, that promote slavery, call for genocide, or incite rape and other forms of violence" as clear examples of cases where normative authority can be justifiably exercised against the democratic will.[44] Yet, even in these "clear" cases, it is difficult to decisively justify what should be done and by whom. Any political decision involving guns in schools, for instance, can be expected to raise complex, quasi-empirical issues related to the welfare of children (and others), and a great many decisions involving labor regulations will raise subtle questions about which status inequalities are morally tolerable. As Peter acknowledges, "the epistemic circumstances of politics tend to be such that [epistemically grounded] normative authority is often difficult to establish."[45] In such an uncertain, risky, and contentious social environment, how could it ever be possible to establish normative authority?

Just as in the context of democratic legitimacy, the linchpin of epistemic legitimacy is social trust and collective autonomy. However, regarding the sort of higher normative authority that is characteristic of epistemic legitimacy, the critical component of collective autonomy is not (collective) authenticity but competence, which is in essence the "ability to effectively form intentions to act, [] along with the various skills that this requires."[46] In most cases, assessing an individual person's competence is usually straightforward: is the person minimally rational, self-controlled, and capable of forming intentions that, under normal circumstances, would be effective? Assessing the competence of a collective, in contrast, is much less straightforward. What would it mean to say that a collective is rational, self-controlled, or capable of forming intentions at all?

The key to understanding collective competence involves seeing that when people act collectively, they often do so through public institutions, formal or otherwise. These institutions can be viewed as population-level tools, whose primary function is to stabilize and govern certain kinds of large-scale civic activity. In the United States, the most effective institutional agencies, such as the National Institutes of

Health (NIH) and the Federal Reserve, embody forms of bureaucratic competence that allow the population as a whole to respond quickly and flexibly to large-scale problems that do not lend themselves to either political or market-based solutions. But, as the history of economic and political development has shown, these institutions cannot be created overnight or imported from elsewhere. To be effective, they must be grown organically over a long period of time, while exhibiting a proven track record of competence. To be credibly viewed as trustworthy, meanwhile, they must be given a degree of independence from mechanisms of direct democratic accountability—such as electoral politics—that is well-matched to their capacities. Under favorable conditions, and only under such conditions, can these institutions serve as truly self-sustaining sources of trust, and insofar as such institutions can manifest forms of collective competence that cannot be obtained otherwise, we will regard them (where they appear) as collectively good in themselves. So, when bad actors sow misinformation to undermine trust in these institutions, without regard to whether they serve a critical role in supporting public infrastructure or providing any sort of alternative, they serve as a drag on a source of epistemic legitimacy.

While collectively aligned democratic assemblies embody democratic legitimacy, effective autonomous bureaucracies embody epistemic legitimacy. As we have argued, both depend crucially on the presence of adequate social trust to function properly. As we will see next, in addition to undermining collective alignment of democratic will, emergent forms of manipulation can also undermine the effectiveness of self-sustaining trustworthy institutions.

3.3. Emergent manipulation and the sources of legitimacy

The practices of CA and the IRA conflict with both democratic and epistemic sources of legitimacy, and without seeming to involve impingements of the autonomy of any particular person.

CA's Do So! campaign bears the hallmarks of emergent manipulation. First, it was stochastic; it did not involve targeting any particular voter for intervention, by getting those specific individuals to behave in any specific way. Rather, the campaign targeted an entire class of voters—youth voters—with the aim of achieving a certain predictable effect only at scale, under specific environmental conditions. Moreover, the campaign did not seek to seriously undermine any one individual's autonomy; that is, exploiting the specific weakness of those who might be highly sensitive to such operations was not the primary goal, and was (in the majority of cases) plausibly not achieved. Rather, the goal was only to persuade a small number of potential voters—recall that Nix described the change as involving only 6% of voters—to feel sufficiently disenfranchised to abstain. Second, the Do So! campaign was fragmented; it did not consist in the open and transparent sponsorship of a political operation. Rather, it involved surreptitious amplification of an existing grassroots movement, paying contributors to propagate the graffiti campaign. Thus, it illicitly borrowed on the populist credentials of that preexisting movement to achieve its goals unencumbered by the mechanisms of accountability that govern political activity.

So, the Do So! campaign falls under the rubric of emergent manipulation. But what--if anything--raises a moral concern with CA's practices in that case? The key threat relates to democratic legitimacy: the practices prevented the political process from reflecting democratic will in the way necessary to avoid collective alienation. While the individual youth voters who abstained from voting might have been able to genuinely affirm their abstention as part of an acceptable autobiographical narrative, the youth voters considered as a group could not have. Indeed, the fact that the Do So! campaign was indifferent to the group's interest in voting, and also depressed that voting, is what makes the campaign manipulative on the definition we articulated in section 2.1.

The IRA campaigns also involve emergent manipulation. Their main mode of operation includes elements of both stochastic and fragmented manipulation. The goal of the active measures was not necessarily to influence any particular individual not to vote (or alternatively, to essentially spoil one's ballot by voting for a third party), but to mix influences with disenfranchising effects into a media ecosystem in which they have the appearance of organically generated content. And as with CA and the Do So! campaign, the primary mechanism by which the IRA exerted its influence was not by wholly disabling the autonomous capacities of any voter, but rather by weakening those capacities or misdirecting them in a subtler fashion. Yet, there is an important difference between the Do So! campaign and the IRA's "active measures" operations, in terms of their effects on legitimacy. The IRA's practices do, of course, threaten democratic legitimacy in many of the same ways as the Do So! campaign did, but the IRA's operations also threaten epistemic legitimacy. They do not aim simply to manipulate persons, either individually or at scale, but they also aim to undermine the legitimacy of institutions that might otherwise serve as self-sustaining sources of trust (and thus, normative authority), such as the independent news media.[47] Without a media that enjoys this sort of trust, a government will not be able to implement and publicly justify policies that are appropriately responsive to reasonable beliefs about what should be done. This problem arises regardless of whether the IRA's operations impinge on individual autonomy, because what is required to avoid this problem is not simply an assembly of individually rational and reasonable citizens, but a citizenry that is holistically unmanipulated, and that shares common knowledge, understanding, and trust.

4. Conclusion

Within any democratic polity, there will inevitably be individuals whose values are unsatisfied, and there will be others who are treated in ways that are alienating. Such individual-level phenomena may threaten legitimacy, but they are not the only threats to legitimacy. And in this chapter, we considered several examples of "emergent" manipulation that operate at the group level, and not necessarily at the individual level. This sort of manipulation, we argued, threatens legitimacy where it is present. In CA's Do So! campaign, the manipulation was stochastic and situated within a polarized and narrowly balanced electoral system where small marginal changes can have a decisive impact. In the IRA's active measures campaigns, the manipulation was stochastic *and* fragmented, in the sense that the interventions were not presented as coming from the IRA, but were distributed to users through multiple, more localized sources of influence. The presence of these forms of manipulation in electoral politics threatens legitimacy. Understanding these forms of emergent manipulation, and avoiding the temptation to understand manipulation and legitimacy as strictly operating at the individual level, we can better understand the range of threats it can present.

References

Alba, Davey. “How Russia’s Troll Farm Is Changing Tactics before the Fall Election.” *The New York Times*, March 29, 2020, sec. Technology.

Ballhaus, Rebecca. “Cambridge Analytica Suspends CEO Alexander Nix after Video’s Release.” *Wall Street Journal*, March 20, 2018, sec. Politics.

BBC Staff. “Cambridge Analytica-Linked Firm ‘Boasted of Poll Interference.’” *BBC News*, March 25, 2018, sec. UK. https://www.bbc.com/news/uk-43528219.

CA Political. “Do So: Trinidad and Tobago,” March 21, 2018. https://ca-political.com/casestudies/casestudytrinidadandtobago2009.

Channel 4. “Exposed: Undercover Secrets of Trump’s Data Firm.” *Channel 4 News*, March 20, 2018. https://www.channel4.com/news/exposed-undercover-secrets-of-donald-trump-data-firm-cambridge-analytica.

Christman, John. *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*. Reissue edition. Cambridge: Cambridge University Press, 2011.

Cohen, Joshua. “Deliberation and Democratic Legitimacy.” In *The Good Polity*, 1989.

Coons, Christian, and Michael Weber, eds. *Manipulation: Theory and Practice*. 1st ed. Oxford; New York: Oxford University Press, 2014.

Detrow, Scott. “What Did Cambridge Analytica Do During The 2016 Election?” All Things Considered, n.d.

DiResta, Renee, Kris Shaffer, Becky Ruppel, David Sullivan, Robert Matney, Ryan Fox, Jonathan Albright, and Ben Johnson. “The Tactics & Tropes of the Internet Research Agency.” *U.S. Senate Documents*, October 1, 2019.

George, Kinnesha. “‘Different Political Beast.’” Trinidad and Tobago Newsday, August 3, 2019. https://newsday.co.tt/2019/08/02/d/.

Hamburger, Tom. “Cruz Campaign Credits Psychological Data and Analytics for Its Rising Success.” *Washington Post*, December 13, 2015, sec. Politics.

Hilder, Paul. “‘They Were Planning on Stealing the Election’: Explosive New Tapes Reveal Cambridge Analytica CEO’s Boasts of Voter Suppression, Manipulation and Bribery.” openDemocracy. Accessed December 28, 2020. https://www.opendemocracy.net/en/dark-money-investigations/they-were-planning-on-stealing-election-explosive-new-tapes-reveal-cambridg/.

Howard, Philip N., Bence Kollanyi, Samantha Bradshaw, and Lisa-Maria Neudert. “Social Media, News and Political Information during the US Election: Was Polarizing Content Concentrated in Swing States?” *Oxford Internet Institute, Project on Computational Propaganda*, 2017.

Kang, Cecilia, and Sheera Frenkel. "'PizzaGate' Conspiracy Theory Thrives Anew in the TikTok Era." *The New York Times*, June 27, 2020, sec. Technology. https://www.nytimes.com/2020/06/27/technology/pizzagate-justin-bieber-qanon-tiktok.html.

Klenk, Michael. "(Online) Manipulation: Sometimes Hidden, Always Careless." *Review of Social Economy*, n.d.

Klenk, Michael, and Jeff Hancock. "Autonomy and Online Manipulation." *Internet Policy Review* 1 (2019): 1–11.

List, Christian, and Philip Pettit. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. 1 edition. Oxford; New York: Oxford University Press, 2013.

Lovett, Adam, and Stefan Riedener. "Group Agents and Moral Status: What Can We Owe to Organizations?" *Canadian Journal of Philosophy*, n.d.

Matz, Sandra C., Michal Kosinski, Gideon Nave, and David J. Stillwell. "Psychological Targeting as an Effective Approach to Digital Mass Persuasion" 114, no. 48 (2017): 12714–19.

Noujaim, Jehane, and Karim Amer. *The Great Hack*. Documentary. Netflix, 2019.

Peter, Fabienne. "The Grounds of Political Legitimacy." *Journal of the American Philosophical Association*, forthcoming.

Pettit, Philip. *Just Freedom: A Moral Compass for a Complex World*. 1st edition. New York: W. W. Norton & Company, 2014.

Premdas, Ralph. "Ethno-Nationalism and Ethnic Dynamics in Trinidad and Tobago: Toward Designing an Inclusivist Form of Governance." In *The Palgrave Handbook of Ethnicity*, edited by Steven Ratuva, 809–24. Singapore: Springer, 2019. https://doi.org/10.1007/978-981-13-2898-5_167.

Rawls, John. "Justice as Fairness: Political Not Metaphysical." *Philosophy & Public Affairs* 14, no. 3 (1985): 223–51.

———. *Political Liberalism*. New York: Columbia University Press, 2005.

Raz, Joseph. *The Morality of Freedom*. Oxford: Oxford University Press, 1988.

Rubel, Alan, Clinton Castro, and Adam Pham. *Algorithms and Autonomy*. Cambridge, United Kingdom ; New York, NY: Cambridge University Press, 2021.

Select Committee on Intelligence, United States Senate. "Report of the Select Committee on Intelligence, United States Senate, on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume II: Russia's Use of Social Media and Additional Views." Washington, D.C: Select Committee on Intelligence, United States Senate, 2019.

Susser, Daniel, Beate Roessler, and Helen Nissenbaum. "Online Manipulation: Hidden Influences in a Digital World." *Georgetown Law Technology Review* 4 (2019): 1–45.

U.S. Department of Justice. “Report on the Investigation into Russian Interference in the 2016 Presidential Election, Volume I (‘Mueller Report’).” Washington, D.C.: U.S. Department of Justice, March 2019.

———. “Report on the Investigation into Russian Interference in the 2016 Presidential Election, Volume II (‘Mueller Report’).” Washington, D.C.: U.S. Department of Justice, March 2019.

U.S. v. Internet Research Agency, LLC, No. 1:18-cr-00032 (U.S. District Court for the District of DC February 16, 2018).

Weaver, Matthew. “Facebook Scandal: I Am Being Used as Scapegoat – Academic Who Mined Data.” *The Guardian*, March 21, 2018, sec. UK news.

Wylie, Christopher. *Mindf*ck: Cambridge Analytica and the Plot to Break America*. New York: Random House, 2019.

Yeung, Karen. “‘Hypernudge’: Big Data as a Mode of Regulation by Design.” *Information, Communication & Society* 20, no. 1 (January 2, 2017): 118–36.

[1] For brevity, we include Cambridge Analytica’s former parent company, the SCL Group, under the simple heading of “Cambridge Analytica.”

[2] U.S. Department of Justice, “Report on the Investigation into Russian Interference in the 2016 Presidential Election, Volumes I & II (‘Mueller Report’)”; Select Committee on Intelligence, United States Senate, “Report on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election,” Volumes I & II; DiResta et al., “The Tactics & Tropes of the Internet Research Agency.”

[3] For the original report, see Channel 4, “Exposed: Undercover Secrets of Trump’s Data Firm.” Alba, “How Russia’s Troll Farm Is Changing Tactics before the Fall Election.”

[4] Channel 4, “Revealed.”

[5] Sources from the *Wall Street Journal* described “Mr. Nix’s penchant for exaggerating the company’s capabilities and work, sometimes to its own detriment.” See Ballhaus, “Cambridge Analytica Suspends CEO Alexander Nix after Video’s Release.” See also Wylie, *Mindf*ck: Cambridge Analytica and the Plot to Break America*.

[6] BBC Staff, “Cambridge Analytica-Linked Firm ‘Boasted of Poll Interference.’”

[7] For a good introduction to the political dynamics of Trinidad and Tobago, see Premdas, “Ethno-Nationalism and Ethnic Dynamics in Trinidad and Tobago.”

[8] Noujaim and Amer, *The Great Hack*.

[9] BBC Staff, “Cambridge Analytica-Linked Firm ‘Boasted of Poll Interference.’”

[10] CA Political, “Do So: Trinidad and Tobago.”

[11] Noujaim and Amer, *The Great Hack*.

[12] George, “‘Different Political Beast.’”

[13] Hilder, “‘They Were Planning on Stealing the Election.’”

[14] U.S. v. Internet Research Agency, LLC, No. 1:18-cr-00032 (U.S. District Court for the District of DC February 16, 2018).

[15] U.S. Department of Justice, “Report on the Investigation into Russian Interference in the 2016 Presidential Election, Volume I (‘Mueller Report’),” 25.

[16] Select Committee on Intelligence, United States Senate, “Report of the Select Committee on Intelligence, United States Senate, on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume II:

Russia’s Use of Social Media and Additional Views” (Washington, D.C: Select Committee on Intelligence, United States Senate, 2019), 6.

[17] Select Committee on Intelligence, United States Senate, 6–7.

[18] U.S. Department of Justice, “Report on the Investigation into Russian Interference in the 2016 Presidential Election, Volume I (‘Mueller Report’),” 24–25.

[19] U.S. Department of Justice, 26.

[20] U.S. Department of Justice, 31–32.

[21] Howard et al., “Social Media, News and Political Information during the US Election: Was Polarizing Content Concentrated in Swing States?,” 1.

[22] DiResta et al., “The Tactics & Tropes of the Internet Research Agency,” 8–10.

[23] DiResta et al., 65–66.

[24] Matthew Weaver, “Facebook Scandal: I Am Being Used as Scapegoat – Academic Who Mined Data,” *The Guardian*, March 21, 2018, sec. UK news.

[25] Tom Hamburger, “Cruz Campaign Credits Psychological Data and Analytics for Its Rising Success,” *Washington Post*, December 13, 2015, sec. Politics.

[26] Scott Detrow, “What Did Cambridge Analytica Do during the 2016 Election?,” *NPR*, accessed June 17, 2020, https://www.npr.org/2018/03/20/595338116/what-did-cambridge-analytica-do-during-the-2016-election.

[28] Matz et al., 12717.

[29] Coons and Weber, *Manipulation: Theory and Practice*; Yeung, “‘Hypernudge’: Big Data as a Mode of Regulation by Design”; Susser, Roessler, and Nissenbaum, “Online Manipulation: Hidden Influences in a Digital World.” In (2021), we argue for an ecumenical account of autonomy, encompassing both agents' relationship to their wills and social structures within which agents' values, understandings, and preferences develop. Autonomy requires both a degree of non-alienation and social structures that foster the ability to develop values, understandings, and preferences within reasonable alternatives. See Rubel, Castro, and Pham (2021), pp. 21-42.

[30] Klenk and Hancock “Autonomy and Online Manipulation.”

[31] Kang and Frenkel, “‘PizzaGate’ Conspiracy Theory Thrives Anew in the TikTok Era.”

[32] See Klenk, “(Online) Manipulation: Sometimes Hidden, Always Careless.”

[33] Peter, “The Grounds of Political Legitimacy.”

[34] Rawls, *Political Liberalism*, chap. 4.

[35] Pettit, *Just Freedom*, chap. 5.

[36] Rawls, “Justice as Fairness: Political Not Metaphysical,” 242.

[37] Peter, “The Grounds of Political Legitimacy.”

[38] Raz, *The Morality of Freedom*, 53.

[39] Cohen, “Deliberation and Democratic Legitimacy.”

[40] Christman, *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*, 154. Marina Oshana, *Personal Autonomy in Society*, pp. 21-49.

[41] Christman, 155.

[42] List and Pettit, *Group Agency*.

[43] Lovett and Riedener, “Group agents and moral status: what can we owe to organizations?”

[44] Peter, “The Grounds of Political Legitimacy.”

[45] Peter, “The Grounds of Political Legitimacy.”

[46] Christman, *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*, 154.

[47] DiResta et al., 65–66.