

Master's thesis submitted in fulfillment of the degree of Master of Applied Ethics
Faculty of Humanities of Utrecht University



Reason for the Digital Age:

Developing Habermasean Tools for Political Moderation in the Post-Truth
Era

Name: Augusto Alberto Santos Ferro

Organizations: Utrecht University

First Reader: Chris Wareham

Second Reader: Hanno Sauer

Place & Date: June 17th, 2022; Utrecht

Word Count: 16,959

Abstract

This thesis aims to introduce a novel approach to the problems of fakenews, political polarization and social media platforms with respect to the ongoing state of the American political discourse through the work of Jürgen Habermas. Habermas' work provides unique insight into the problems of communication and action coordination which are of utmost relevance to this discussion. This paper argues that social media platforms are formulated around a model which is highly vulnerable to the influences of fakenews content, and therefore instrumental reason. Furthermore, this thesis continues on to argue that this reliance on instrumental reason is not an intrinsic feature of these networks, but rather, that a platform could indeed be designed which embraces communicative action as its primary avenue of social cooperation. In so, the problem of polarization within the American political discourse might be ameliorated through the exercise of communicative rationality if provided a sufficiently rich argumentative space.

This thesis offers in first place, a Habermasean critique of the inner workings of social media platforms with respect to their vulnerability to instrumental reason. Thereafter, continues on to articulate specific features which would be of greatest impact with respect to the cultivation of communicative rationality on these networks.

Table of Contents

<u>Abstract:</u>	
<u>Section (I): Introduction</u>	3
<u>Section (II): Fakenews and the Problem of Political Polarization</u>	6
i) Fakenews in a Nutshell: Digital Era Propaganda	6
ii) How does Fakenews Operate?	10
iii) Avenues of Action: Censure versus Critical Thinking	13
<u>Section (III): Political Polarization & Rising Social Conflictedness in the Contemporary United States</u>	15
i) Social Media, Fakenews & Political Polarization in the United States	15
ii) Republicans and Democrats, a Widening Ideological Gap	17
iii) Social Conflictivity in the Political Discourse of the United States	19
iv) Recent History of Massive Civil Unrest	20
v) Wearing Rose Colored Glasses: Post-Truth and the Mediation of Facts by Beliefs	21
<u>Section (IV): A Theory of Communicative Action - Key Concepts</u>	23
i) Theoretical Underpinnings: Austinean Speech Act Theory	24
ii) Habermas' Argumentation Theory:	26
iii) Instrumental Reason & System	29
iv) Communicative Action & Lifeworld	31
<u>Section V: A Habermasean Approach to the Problem at Hand</u>	33
i) Divergent Lifeworlds & the Abandonment of Communicative Rationality	34
ii) A Habermasean Solution to a Digital Age Problem	36
<u>Section VI: Speaking to Each Other, Rather than @eachother</u>	37
i) Replacing Forum Posting: Cultivating an Argumentative Space	38
ii) Replacing Comments: Normative Regulation of Argumentative Speech Acts	40
iii) Dissolving Anonymity: Engendering Accountability between Peers	41
iv) Speaking to Each Other Rather than @eachother	41
<u>Section VII: Risks Related to an Intersubjective Formulation of Social Media Networks</u>	42
<u>Section VIII: Concluding Remarks</u>	44
<u>References</u>	47

Section (I): Introduction

Paired with the rise of the 21st century is the emergence of digital communications and social media. New technologies have allowed people to communicate across vast distances with each other, apparently in an instant. The digital medium has brought us all closer together, bridging what are otherwise insurmountable physical impediments, allowing people to communicate with each other irrespective of where or when they are voicing themselves. Despite promoting interpersonal discourse across the world between otherwise disconnected people; is this increased connection and closeness beneficial to our society? Or perhaps, does it act as a polarizing force? Has it truly pushed us together into a more fraternal society? Or has it merely sat us into a pressure cooker with our social rivals?

The advent of social media networks at the beginning of the century represents a wholesale projection of our own social practices onto this newfound medium; a platform in which ordinary social interactions might be reproduced in the digital world. It provides a virtual space wherein people can communicate with each other, transferring not only text but any type of multimedia files with the greatest of ease. Today, this virtual space has become all-encompassing, a crucial tool for any person who aims to fully participate in society. Not only do social media networks facilitate communication between individuals, but they oftentimes mediate one's relationship to both commercial and governmental entities. Social media platforms house the greater constellation of public and private actors which are constitutive of society at large, whether they be individuals or groups. Everybody participates in the discourse which occurs on social media networks in some way or another, they are the primary instrument of social interaction today.

These platforms operate under the premise of unifying people across the world through the increased facility of interpersonal communication they provide. In their mission statement, Meta (formerly Facebook) for example, aims to “to give people the power to build community and bring the world closer together” (Meta, 2022). The sentiment behind this goal is shared widely across all social media networks, it is explicitly their *raison d'être*. However, to what extent is this goal achieved by these platforms broadly?

Spurring interpersonal communication can have its downsides however, in the late 2010s we witnessed the emergence of ‘fakenews’, a term first coined by former president of the United

States, Donald Trump. This term initially indicated totally false news content, and carried with it the implication that the content should be outright rejected. Today however, the term has become a political weapon which people use to bludgeon news content and opinions which they do not believe to be correct. To what extent has the phenomena of fakenews fueled the recent polarization of political discourse?

Recent studies would lead us to believe that we live in an evermore polarized world, one in which people's worldviews are becoming incommensurable (Pew, 2014; Hong & Kim, 2016; Connaughton, 2021). Moreover, any person with access to Twitter for example, might also affirm this same statement given a typical user experience. In this context, the US seems to bear the brunt of this polarization; their political system being strongly affected by the rise of these newly incommensurate worldviews. Anecdotally, many political commentators have observed that the political divide in the US is reaching never before seen heights of incompatibility. To what extent might this newfound political divergence be rooted in the ways in which we communicate our political ideas on these social media platforms?

According to the work of thinkers like J. Habermas, perhaps the structure of social media itself promotes the deepening of divides between relevant social actors and their worldviews. In his work *The Theory of Communicative Action*, Habermas provides an ardent critique of instrumental reason, instead, offering a model of communicative action as its replacement. The latter focuses on promoting honest discourse via intersubjective elucidation of the world, as opposed to the former model, which is grounded on means, ends, and objectification. Habermas provides an interesting perspective into this problem given his fixation on the conventional consequences of language use, and more importantly, the coordination of action within rational communities. Together, these two features are able to provide a novel perspective with respect to how the proliferation of fakenews propaganda affects the ongoing polarization of the political discourse in the US.

This thesis aims to investigate the link between fakenews propaganda and political polarization in the US through the work of Habermas. It asks the fundamental question, what features of social media networks facilitate the spread of fakenews and therefore the polarization of political discourse? Moreover, could these features be reformulated in a way which instead impedes the propagation of fakenews content, thereby cultivating moderation in political discourse?

This thesis argues that social media platforms are highly vulnerable to the spreading and sharing of fakenews content due to their instrumental formulation, and therefore contribute to the radicalization and polarization process of today's political discourse. On the other hand, this paper argues that this reliance on instrumental reason is not an intrinsic feature of these networks, instead; a platform which is constructed on the basis of a Habermasian approach, grounded on a model of communicative action, could work to effectively bridge the ever widening divide in political ideation which has emerged in the contemporary US.

I argue that social media platforms are designed in a way which facilitates the sharing and spreading of fakenews style propaganda due to their vulnerability to instrumental reason and action, thus, in turn provoking a polarization of the real world political discourse. This style of digital era propaganda contaminates much of the discourse which occurs on these platforms, effectively tainting any possibility of productive or unifying discussion between users, slowly but surely rendering their worldviews wholly incompatible with that of their perceived political opponents. These platforms inadvertently push this type of propaganda due to its particular formulation inherent to the style of fakenews propaganda. From this standpoint, we can observe the emergence of 'echo chambers' or 'bubbles' wherein individuals become more and more radicalized due to a lack of opposing views within their social circles.

Furthermore, I argue that a social media platform could be designed around Habermas' conception of communicative action which might work to slowly ameliorate the issue of increased political polarization. Rather than any paradigm of instrumental reasoning which is characteristic of the design of traditional social media networks, such a platform, if formulated correctly, would strive to promote the use of communicative rationality as the primary vehicle of a person's online discourse.

In order to achieve these goals, section (II) of this thesis will briefly outline the problem of fakenews and how it operates through social media platforms. Following this, section (III) will examine the precarious nature of US political discourse in relation to the ideological consistency and social conflictedness of American society, used as an exemplary case of increasing polarization. Section (IV) will address crucial concepts which must be drawn upon from the work of Habermas in order to later assess the failures of social media platforms insofar that they incentivize the spreading of polarizing content in the vein of fakenews. Section (V) will provide a novel approach to the problem of fakenews and online radicalization through the lens of

Habermas' work. Section (VI) will provide an alternative formula for the construction of a social media platform, one which aims to incentivize communicative rationality between its users. Section (VII) will address various critiques of such a system with respect to both its efficacy and potential risk for increased polarization. Finally, section (VIII) will provide some concluding remarks with regards to the problem of communicative action and social media networks.

Section (II): *Fakenews and the Problem of Political Polarization*

Characteristic of the late 2010s, the concept of fakenews which emerged under the Trump administration is an unignorable historical occurrence. This newfound stylization of the term 'propaganda' was born and raised on social media networks. Not so long ago, Americans would gather *en masse* to shout the term at journalists during Trump's political rallies. Today, fakenews has taken a different form, instead of merely being used for its accusatory value, now it is pushed and trafficked across the web in an effort to persuade masses of people to take direct action. Used by both left wing and right wing movements, fakenews aims to mobilize the masses to take what is oftentimes violent action towards the achievement of a certain political goal. Whether it take the form of an outright lie, subtle obfuscation of the truth, or purposeful misinterpretation of fact; fakenews content is there to persuade people to take up arms in support of a specific cause.

The following section will continue on to address the characteristic features which differentiate fakenews from traditional forms of propaganda in its first subsection (II, i). Thereafter, the modus operandi of fakenews content will be illustrated in subsection (II, ii). Finally, the two fundamental avenues of control in relation to the proliferation of fakenews content will be evaluated in terms of their consequences in subsection (II, iii).

i) Fakenews in a Nutshell: Digital Era Propaganda

We can understand fakenews to be any sort of propagandistic content which is designed to motivate political discourse or action through the means of mass communication. This sort of content can be used to agitate or incentivize people to take action for the sake of a particular social or political movement. We can easily divide the phenomena into two broad categories of fakenews according to researchers: misinformation and disinformation. Misinformation referring

to content which is “bogus or deceptive”; and disinformation referring to “inaccurate information that is deliberately distributed to mislead readers” (Ralston et al, 2018; 31). The former category refers mostly to that content which is shared under the presumption of truth, but in reality, is propagandistic in nature. The latter category refers to that content which is intentionally designed to deceive or misdirect its reader away from the reality of a situation. Together, these propagandistic tools successfully taint the political reality of vulnerable individuals, leading them to participate in the realization of the social or political goals of a third party.

Beyond merely this definition, fakenews carries with it three crucial features which make it so virulent. Firstly is its decentralized model of production, which intends that consumers also participate in the creation and dissemination of content. This participatory model provides a great degree of flexibility and adaptiveness to fakenews content, allowing it to effectively adapt to shifting circumstances and survive contradiction. Secondly, of note is the permeability and extensiveness of the term fakenews, commonly used to indicate any type of factual error. Despite offering an apparently critical attitude, this feature aims to foment the perception of factlessness, potentially making all factual content vulnerable to the accusation of fake news. Finally, the accusatory value of the term cannot be ignored, more often than not, such an allegation is used to mark certain content for wholesale rejection by the ideological community. This functions to reduce the plausibility of content which aims to contradict the narrative constructed by fakenews narrative a priori to its audience’s consumption.

Firstly, is its decentralized model of production and dissemination. Whereas in the 20th century, one would need to persuade or bribe certain journalists to have a story inserted in a national newspaper. The 21st century has seen the massification of our news-bearing institutions. Fakenews thrives on the decentralized model of social media platforms, being pushed and trafficked by unwitting users who lack the research infrastructure to confirm claims before publishing them. Studies are beginning to demonstrate the prolificness of this mass hawking of fakenews content, oftentimes being packaged in the form of memes or short posts. A study which investigated the online interactions between politically minded communities on Instagram was conducted by researchers in 2021 to this respect. Here, research was collected regarding roughly 550,000 posts sent between nearly 200,000 users between the years of 2012 and 2018 with the aim of evaluating how online communities discuss news content (Al-Rawi, 2021; 1). The author of the study readily affirms that fake news discourses found on the platform indicate

the existence of two highly polarized factions which are constituted by opposing political allegiance, furthermore that these communities readily use memes for the sake of attacking and demeaning their opponents (Al-Rawi, 2021; 12). These factions, characterized by their competing political allegiance, traffick political discourse through the means of memes and image-based posts on Instagram, thereby crowdsourcing the construction of fakenews narratives on a partisan basis. The researcher goes as far as to characterize the gravity of the situation by asserting that “Meme War II is a daily reality on Instagram” (Al-Rawi, 2021;12). Through the interpretation provided by this study, we can readily observe how fakenews content is couched in a certain conception of warfare, it is weaponized and utilized as one of many tools in a propagandist’s arsenal. This model of participation continually transforms consumers into creators and vice versa. In so, it is not so easy to identify concrete sources of this style of propaganda due to its decentralized model, but rather, entire online communities have become the publishing mill of these scantily evidenced claims. Furthermore, being that its consumers participate in its production process, fakenews can be highly adaptive with regards to problematic circumstances which might emerge as an affront to its veracity (such as unfulfilled predictions). Through this model of propagation, material inconsistencies identified by the audience naturally correct themselves in the next iteration of the content, an ideological arms race whereby either faction readily upgrades their claims in reference to newfound defects. In so, fakenews content is continually being refined and polished by a plethora of faithful followers, growing ever more resistant to counter-narratives. In so, fakenews content has proven to hold a great degree of longevity in the public sphere, only being temporarily vulnerable to counterclaims. This decentralized model of production and dissemination enhances the perceived reality around their propagandistic content via the creation of an ideological community; a direct consequence of this feature being the increased flexibility and longevity of fakenews content on online platforms.

Another crucial feature of fakenews content is its permeability and extensiveness. Although once used to refer to wholly false content, along the lines of disinformation; today we would do better to characterize fakenews in the terms of misinformation. Indeed, fakenews is used to indicate the failing veracity of any factual claim; and any complex of claims is bound to have a bad apple in it. This fact implies that all content which asserts claims about an ongoing state of affairs is vulnerable to the accusation of fakenews. The term can be applied to almost any

informative content, whether it be posted online or published on traditional media. Whether a piece of content be wholly false or misleading, or simply include a small proportion of false claims; both may be qualified under the category of fakenews. Around the world, but most specifically in the US, organizations such as Politifact have emerged since the late 2000s with the aim of qualifying both individual user posts and institutionally published content along a qualitative metric represented by its patented “Truth-O-Meter”. Oftentimes referenced as a reliable metric in American political discourse, Politifact’s measurement exemplifies the pervasiveness of fakenews content, it can be trafficked through any seemingly factual claim in a more or less significant way. In recent years, politifact’s tool has been cited frequently in news and media reports as a relatively objective measure with regards to the factuality of claims.

Finally, such a discussion would be incomplete by merely addressing the content of fakenews, most importantly is the term’s accusatory function. Not only does fakenews serve as a vehicle for the transmission of dubious information, it also serves to mark content which should be outright rejected by the consumer. To the audience of a fakenews propagandist, the declaration of fakenews implies that the contrarian content in question has been qualified as taboo, and deserves uncritical rejection by their faithful followers. This feature is best exemplified through former president Trump’s use of the term throughout his tenure in office. Oftentimes, he would berate journalists who were present at his political rallies, repeatedly accusing them of pushing fakenews propaganda. In fact, this became one of the major themes of his political platform, resisting the established news media infrastructure through outright rejection. In 2020, research was conducted which tracked how the former president would address the news media over the course of his first year in office. The author of the study concluded that “phrases [such as fakenews] signaled to the public that this entity was no longer worthy of Trump’s or the public’s time” (Meeks, 2020; 230). This accusatory function of fakenews serves to inoculate its consumers against contrarian claims or divergent perspectives. These counterclaims are naturally dismissable given the authoritative declaration given by the content’s creator; uncritically rejected as heresy. The study goes on to assert that using the term fakenews in such a fashion serves to foment “an increasingly polarized political climate and fosters a tribalism ... around sources of information and truth” (Meeks, 2020; 231). This assertion should be an unsurprising one given the accusatory force behind the term. The fakenews propagandist is able to create social communities which fanatically adhere to their own

division of true and false content, thereby creating an ideologically resilient group of followers, reinforced against anything which might aim to contradict their own narrative.

The effectiveness of these features can be observed through the radicalization of the Q-Anon conspiracy theorists. These ideological radicals possessed an unyielding belief that the former president was the true victor of the 2020 election, and was imminently going to recapture the US government via a populist coup d'état. This conspiracy theory proved to be incredibly flexible, adapting to many discrepancies which emerged between the theory's prophetic narrative and the real world events which actually transpired. This fakenews narrative demonstrated a great degree of longevity; with supporters continuing to follow the conspiracy well after Biden's investiture as president. More specifically, the Q-Anon conspiracy is an exemplary case of how the virality of fakenews content can lead to real world polarization of political discourse; to the extent that it led to historically significant incidents such as the breaching of the US Capitol by rioters in 2021.

In summation, fakenews aims to degrade someone's relationship to truth and fact by supplanting one's belief structure on one hand, and by delegitimizing contrarian views on the other. We would do well to consider fakenews any piece of content which masquerades as some sort of legitimate content, either intentionally or unintentionally, while really being a work of either disinformation or misinformation. Further, this type of content is formulated in such a way that it is easily disseminated across broad swaths of social media networks, seeking out those most vulnerable to persuasion. Finally, no source of information is free from the grasp of fakenews accusations; it can be used to qualify any piece of content which is shared across society's information-bearing institutions. In its most minimalistic sense, fakenews is the counterfactual milieu which is an outgrowth of any factual claim; in its most radical sense we might consider it wholly fabricated information, designed for the purposes of coercing vulnerable individuals towards the achievement of specific political goals.

ii) How does Fakenews Operate?

Fakenews operates in plain sight, more often than not hiding itself within true or legitimate content. However, fakenews content thrives due to three specific social factors: confirmation bias, algorithmic selection, and online echo chambers. Firstly is confirmation bias,

which can be understood as a psychological tendency of people in which individuals who are confronted by factual claims which contradict their fundamental belief structure will oftentimes reject them, instead, choosing to reinforce their counterfactual belief system in response. Next is the strong algorithmic preference which is programmed into platforms such as Twitter for the purpose of effectively disseminating this content. Finally, the greatest factor which facilitates the spreading of fakenews is the congregation of ideologically homogeneous groups, which publish, share, and relay content to each other known as echo chambers. These groups are highly vulnerable to the spread of fakenews because of their lack of ideological diversity, and therefore lack of critical motivation when consuming such content. Together these three factors work in conjunction to provide an environment which is ripe for the production, dissemination, and consumption of fakenews content.

The first step in creating an environment which facilitates the creation and distribution of such content is addressing people's psychological tendency to reinforce their belief systems when confronted by factual counterclaims: confirmation bias. A commonplace concept in today's culture, confirmation bias is "typically used in the psychological literature, connotes the seeking or interpreting of evidence in ways that are partial to existing beliefs [or] expectations" (Nickerson, 1998; 175). This psychological tendency leads to individuals discounting factual counterclaims on the basis of their disruptive force with respect to a preexisting belief system. Such a situation could be exemplified by a mother who refuses to accept her son's delinquency when confronted by police officers: 'my son is a good boy, he would never do that!'. Translated into the terms of fakenews and online propaganda, people reject information which might contradict their political beliefs. Digital propagandists utilize this tendency to their advantage, tailoring content to the political views of a specific audience. In so, they successfully avoid many prickly counterclaims which would upset their audience's belief.

This opportunistic targeting of people's confirmation bias is further enhanced by the algorithmic selection of content common to all social media platforms. In 2021, the largest systematic study concerning Twitter use and algorithmic selection was conducted by Huszár et al; these researchers aimed to investigate the extent to which the platform was responsible for the dissemination of extreme and/or highly partisan content. To do so, they performed a comparative analysis of the content recommended by the platform across a number of countries. Despite not evaluating fakenews itself, Huszár et al's criterion of 'highly partisan' content is sufficiently

similar to our own definition of disinformative fakenews to be of use: namely, that it be “deliberately distributed to mislead readers” (Ralston et al, 2018; 31) and could be considered to be an obfuscation of the truth. Being that highly partisan content aims to achieve one party’s electoral goals through the one-sided presentation of their policies and rhetoric; we can also claim that it is deliberately designed to mislead its readers with respect to the legitimacy of other partisan positions via omission, pretending to be the ‘only’ or ‘best’ option. Despite finding that Twitter’s algorithmic selection of content did not incentivize extremist views over that of ordinary political perspectives; researchers did find that the platform strongly incentivized news stories with a heavy partisan bias (Huszár et al, 2021). This algorithmic bias fuels the spread of fakenews content. As suggested by Huszár et al, Twitter holds a preference for feeding partisan content to a partisan audience; those who are most vulnerable to the unwitting consumption of fakenews. Furthermore, this preference naturally serves to incentivize the creation of echo chambers, facilitating the construction of ideologically homogeneous groups online due to the lack of contradictory claims presented to their readers on a personalized basis. In so, Twitter inadvertently creates a multiplicity of incommensurate worldviews, each grounded on a dissimilar notion of what is and is not fact, ultimately constituted by belief in those ‘facts’. By presenting different people with different content, more importantly, content which is tailored to their political allegiance, Twitter cannot help but deepen ideological divides. In so, the average twitter user is put into the position to believe that their perspective is not only true, but widely held by society at large, even if it is untrue amongst their real world social peers.

Together, these two features of the fakenews propagandist’s strategy cultivate fertile grounds for the emergence of their preferred operational environment: echo chambers. In his 2020 publication, P. Barberá characterizes echo chambers as a social space “where citizens do not see or hear a wide range of topics or ideas, which limits their capacity to reach common ground on political issues” (Barberá, 2020; 38). These spaces congregate people into ideologically homogeneous social circles, where they are unlikely to encounter contrarian perspectives or ideologically divergent narratives. In so, individuals who have already been made vulnerable to fakenews content, are transformed into a captive audience. These echo chambers are constituted of a specific set of beliefs or facts which are taken as dogma, and are thereby uncritically accepted. The construction of social echo chambers incentivizes the ideological isolation of captured individuals, further precluding any critical thinking with respect to the

fakenews narrative being pushed. In such an environment, propagandistic content thrives: a social space wherein rational counterclaims might be outright dismissed via the normatively regulated structure of the community at large, simply given their failure to adhere to the group's prior belief system. Such a failure can only result in one outcome, the a priori rejection of counterclaims. In so, echo chambers are utilized by digital propagandists as a self-regulating instrument, one which aims to create and expand ideologically homogeneous groups; those which are most vulnerable to the proliferation of fakenews content.

Fakenews content plays on well studied psychological tendencies which are innate to human reasoning. By taking advantage of people's proclivity to reinforce their own views before dismantling them, digital propagandists are able to foment the creation of echo chambers. Furthermore, social media platforms like Twitter inadvertently fuel the growth of these closed groups via the algorithmic selection of content which is presented to its users. Together, both confirmation bias and algorithmic selection serve as the fakenews propagandist's primary tools for cultivating their preferred hunting grounds: ideologically homogeneous groups, or echo chambers.

iii) Avenues of Action: Censure versus Critical Thinking

Naturally, the question which arises when confronted by the propensity of fakenews on today's social media networks is how could the spread of this content be impeded? The solution to this problem can be approached from two primary vectors; from above or below. Either some centralized entity should be invested with the powers to purge these networks of this content as to protect those most vulnerable, or; that consumers of this content should become more resilient to such malignant efforts of persuasion.

The former solution, censorship, is the traditional avenue of action. Going back to even the times of the ancient Greeks, censure has been considered an effective tool for silencing spreaders of misinformation the likes of Socrates for millennia. Today however, censorship is a much more lenient process. In the terms of online censure, social media platforms compare any given post to a series of normative conditions which determine what is and is not appropriate for the platform. These conditions usually track alongside our own social standards for comportment quite well; oftentimes concerned calls to violence, vulgarity, or profanity. Once identified,

censure-worthy content is deleted, and its uploader warned about further posting of such content; these warnings eventually lead to an outright ban from the platform. In so, problematic content is removed, and problematic people along with them.

On the other hand, we could look towards reinforcing the ability that the victims of fakenews have to defend themselves. From this perspective, we are taking an approach which teaches individuals how to resist propagandistic content via enhanced identification and resilience with respect to their consumption of fakenews. Through educational efforts, great strides could be made in teaching people on how to confront fakenews, namely, through reinforcing their research skills and by offering quick guides for identifying commonplace forms of misleading content. Through this method, the need for censure might evaporate; in an ideal world, people would be made intellectually resilient enough to resist the persuasive influence of propaganda itself.

Neither solution provides a clear pathway towards the eradication of fakenews online. The latter solution requires an immense amount of time and effort, moreover, implies a heap of collateral damage due to the approach's sluggishness. Perhaps this solution indeed relies too heavily on the intellectual virtue of those people who are already vulnerable to manipulation and persuasion from malignant actors. On the other hand, the former solution reeks of heavy handed authoritarianism and means-ends thinking. Furthermore, studies are beginning to show that censorship may in fact reinforce the radical views of those censored. According to some researchers, users which are banned for expressing a specific belief will oftentimes seek out others who believe similarly, ultimately resulting in the creation and growth of radicalized communities elsewhere online (Lane et al, 2021; 13). This unfortunate reality of censorship implies that it might do more harm than good; simply excising people from digital society, without impacting their radical views. Despite accepting a certain amount of censure for the common good, when ought to be very mindful of the real world consequences of such measures. Of note is the implication of numerous unfavorable moralistic and pragmatic consequences of the former, while the latter aims to reinforce the intellectual resilience of those who are susceptible to fakenews content.

Section (III): Political Polarization & Rising Social Conflictedness in the Contemporary United States

The problem of political polarization and civil unrest in the US is one which has become painfully evident in the past decade. In the past five years the world has witnessed mass outbreaks of violence emanating from both left and right wing political movements. From the George Floyd riots which erupted throughout the summer of 2020, to the Capitol Riots of January 6th, the US is suffering an unprecedented degree of civil unrest. Without regard for ideological justification, these incidents can be fundamentally characterized as impassioned expressions of political dissent in relation to some socio-political trigger issue. In the case of the former, the trigger being police reform, and electoral integrity in the case of the latter. Furthermore, other powder keg issues are quickly appearing on the horizon, namely the recent supreme court decision to roll back federal abortion protections as a consequence of a reinterpretation of the Roe v. Wade decision of 1973, which is set to take effect over the coming summer. What factors might be contributing to this deepening divide in the US?

To further explain this issue, the remainder of this section will be divided into five parts. Firstly, in subsection (III, i), the relationship between social media platforms and political polarization in the US legislature will be investigated through a 2014 study conducted by Hong & Kim. Thereafter, data drawn from Pew Research Center will be reviewed with the intention of evaluating the current state of political radicalization and polarization in the US in subsection (III, ii). Following this, another study published by Pew Research will be used to illustrate the growing trend of civil unrest in the country in subsection (III, iii). Then, the recent history of US civil unrest in the past decade will be briefly reviewed in subsection (III, iv). Finally, in subsection (III, v), the four previous sections will be evaluated through the lens of M. Comacho's characterization of the "post-truth era" in his 2019 article. Together, these studies would lead us to believe that Comacho has far from mischaracterized the ongoing crisis in political discourse which currently exists in the US.

i) Social Media, Fakenews & Political Polarization in the United States

Fakenews, which has been trafficked across social media networks at ever increasing rates is designed to influence individuals and achieve political goals, and it oftentimes quite successful in its aims. Not only does fakenews succeed at motivating individuals towards mass-scale political action; we can also observe that it achieves its goals inasmuch material changes in legislative policy are concerned. Despite focusing its efforts towards those vulnerable to persuasion, the influence of fakenews eventually flows upwards towards powerful elected representatives as well. In a 2016 study, researchers at Yonsei University investigated the link between radical communities rooted in social media platforms, namely Twitter, and their relationship to elected officials in the US House of Representatives.

This 2016 study aimed to compare the ideology of elected representatives given their legislative activities, to the proportion of Twitter followers they had which could be considered politically radical. Despite admitting that social media is not the primary driver of political polarization in the US, but rather, such a process was most certainly already underway. The authors do claim that due to the highly personalized nature of Twitter's algorithmic presentation of content, this platform does make a noticeable contribution to the increasing polarization of political discourse online (Hong & Kim, 2016; 5-6). Furthermore, the researchers are able to establish a link between radically-minded elected representatives and their number of radical Twitter followers. These researchers take this statistical correlation to imply that government officials should be wary of interpreting social media commentary as an accurate depiction of their constituent's opinions (Hong & Kim, 2016; 6). This warning implies that elected representatives should be mindful of the fact that radicals are oftentimes louder than average constituents, drawing more attention than ordinary people, especially via such an open medium. Moreover, this implies that those elected officials who maintain the most ideologically radical views might be motivated in doing so because of the disproportionately high number of political radicals which follow them online. Taken to its next logical step, if both sides of the aisle are being pulled towards more radical political positions; then the end result is an increase in political polarization and fundamental disagreement between elected officials. Furthermore, we can begin to observe how this dynamic can lend itself to a positive feedback loop of polarization; wherein radicals express more of their views online, elected officials take note and mimic their positions, leading to the radicalization of their own political followers and so on. In general, the study concluded "that political polarization may be especially problematic in social media"

(Hong & Kim, 2015; 6) due to the algorithmic selection of user content. In so, the relationship between fakenews, social media and political polarization can be observed across all levels of society; that it flows upwards and can provoke material changes in governmental policy.

Hong's & Kim's study into how social media platforms can drive real world political polarization is a crucial piece of evidence in observing the process of radicalization online. This study shows how, via social media platforms (those pathways through which fakenews flows most effectively), radically-minded individuals are able to influence political discourse on the highest levels, creating a feedback loop which provokes further radicalization online. In so, more echo chambers are engendered between progressively radicalized groups, each trapped in a prison of confirmation bias. These bubbles therefore, drive the polarization process online, which in turn polarizes the real world political discourse on all levels, bearing weight even onto the activities of elected representatives.

ii) *Republicans and Democrats, a Widening Ideological Gap*

Pew Research Center, founded in 2004, is a renowned American think tank, specializing in political analysis and the interpretation of mass-scale polling data. In the past 10 years, Pew has published various studies of note with respect to the topic of political radicalization in the country. Of greatest concern for the purposes of this thesis, Pew published a study in 2014 which aims to track ideological consistency amongst Americans since 1994. This study, which will soon be updated (presumably by 2024), reveals a shocking trend in the average political values held by Republicans and Democrats respectively. Namely, that the average Republican or Democrat disagrees with each other more deeply, and on a greater share of political positions than ever before. To illustrate these facts, two pieces of data are crucial: the greatness of the gap, and across which parts of society does this gap apply?.

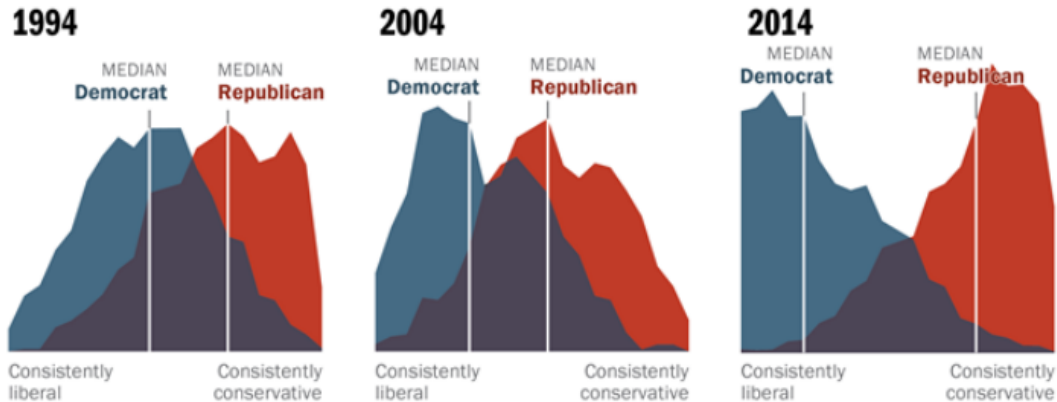
In the case of the former, [in 2014] the ideological gap has never been wider. Critically, this divide is most striking amongst the politically engaged. Figure 1 represents this ideological divide graphically, with Democrats in blue and Republicans in red. As one can naturally observe, the median voter's political values have drifted away from each other quite drastically since 1994. This effect is frighteningly apparent in the crucially relevant category of politically engaged citizens, where the median voter on either side shares nearly no political values with

their opponent. Worse yet, the data which is represented in this graph was compiled years before some of the greatest instances of civil unrest seen in the past few decades.

Polarization Surges Among the Politically Engaged

Distribution of Democrats and Republicans on a 10-item scale of political values, by level of political engagement

Among the politically engaged



Among the less engaged

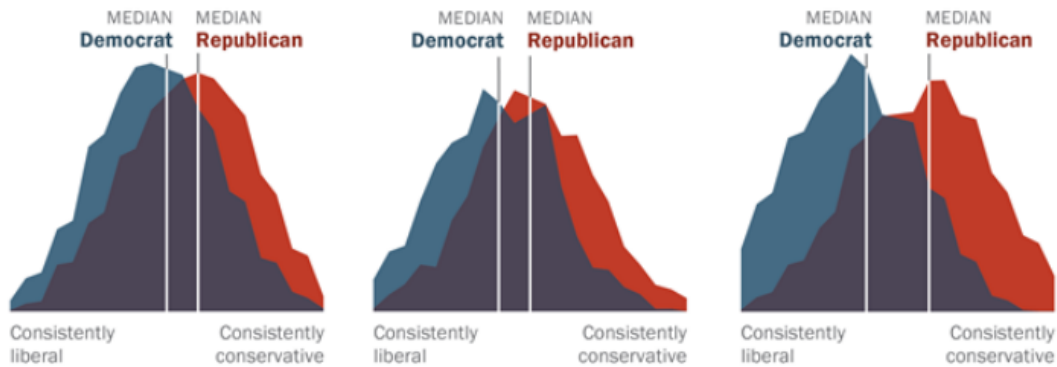


Figure 1 - (Pew Research Center, 2014)

Pew’s own evaluation of the data itself concurs with this interpretation. According to Pew, “fully 99% of republicans are now more conservative than the median Democrat, while 98% of Democrats are more liberal than the median Republican” (Pew Research Center, 2014). This data indicates the absolutely stunning level of polarization; the average voter of either party shares little to no political values with their opponents. Furthermore, “70% of the politically engaged now take positions that are mostly or consistently in line with the ideological bend of their party” (Pew Research Center, 2014). Not only are the political values of Republicans and Democrats almost completely malaligned; but a vast majority of engaged voters appear to be motivated along purely partisan lines. Together, these two facts highlight the great degree of political polarization which exists in the US; specifically amongst those engaged in the political

process. Not only are their values incommensurate with each other, but the political discourse seems to be anchored to the gridlock ridden bipartisan system which characterizes the US legislature.

This 2014 study shows a worrying trend in the world of US political discourse. Since the 1990s the gap between Republican and Democratic voters has been growing wider and wider. In 2014, this gap seemed to be insurmountable, with a vast majority of party members holding incompatible views with that of their political opponents. Since 2014, horrendous demonstrations of civil unrest have erupted across the US from both sides of the political aisle; what might this graph look like in 2024? Overdue for an update, this study should include more recent data in the coming years; hopefully shining some light onto what role the Trump administration may have played in the political polarization of average citizens.

iii) Social Conflictivity in the Political Discourse of the United States

In 2021, researchers operating in conjunction with Pew Research Center conducted another study with respect to social conflictedness in the US compared to other advanced economies. This study aims to highlight the perception of social conflict across various demographic groups. To assess the degree to which the US might be subject to social conflictivity, a number of polls were conducted both nationally and internationally. The two factors tested for in this study are the frequency in which the subject perceives social conflict; and the degree to which participants might concur with the factual claims of their compatriots. Both of these data points fall in line precisely with the notion that the US may be undergoing what many observers have called a ‘post-truth’ period in relation to its political discourse

With respect to the regularity of social conflict, that is, wholesale disagreement between contrasting demographics; Pew has frightening data. According to the research center, “an equal share (90%) of partisans on both sides say there are strong or very strong conflicts between people who support different political parties” (Connaughton, 2021). On the contrary, there appears to be “an overall median of 50% [of people who would] say the same across all advanced economies” (Connaughton, 2021). This data is unsettling for anybody concerned with healthy political discourse. It would appear that the US suffers from nearly double the social conflictedness with respect to its political sphere compared to other advanced economies.

However, it should be noted that doubling the international score would imply total conflictedness according to Pew; the US is incapable of increasing its level of social conflict very much more.

Furthermore, in this study, Pew has also recorded data which tracks whether or not Americans perceive each other to be in agreement with respect to basic facts. Pew affirms that “a majority of Americans (59%) say people can’t agree on [the] basic facts” (Connaughton, 2021) of a political discussion. Furthermore, it appears that “partisans in the US are roughly equally likely to say that Americans disagree about basic facts” (Connaughton, 2021) of a discussion. The problem of ‘basic facts’ in the US is one which cuts across party lines; it does not appear to be a product of either party’s conduct or rhetoric. Pew recognizes that the US is not unique in this level of fundamental disagreement, but rather holds “second [place] only to France (61%)” (Connaughton, 2021). Despite the French people winning out on this metric; this degree of fundamental disagreement does not indicate strong possibilities for social cohesion in the US. Rather, it seems that both countries have undergone periods of dramatic social conflict since the publication of this study in 2021.

This study provides a chilling perspective into the state of US political discourse. It seems as though the US is undergoing a period of extreme political tension. It appears as though both sides of the political debate are incompatible with each other; moreover, that Americans fundamentally disagree with each other. This data becomes even more striking as one pits it against studies done abroad. The US is currently experiencing an unprecedented rupture in its political discourse; Republicans and Democrats are slowly becoming couched into wholly incompatible worldviews.

iv) Recent History of Massive Civil Unrest

The recent history of social conflict in the US should present itself as the most glaring piece of evidence with respect to this newfound level of ideological divergence amongst Americans. Periods in which mass outbreaks of civil unrest have been widespread in the US. The George Floyd riots of 2020 represent an emotive breach of civil peace; violent confrontations between American citizens, police, and other citizens became commonplace across the country. On the other hand, the US capitol was penetrated by a group of rioters in 2021, whose goal was

the feverish reconquest of the American government. In both examples, social media platforms and fakenews were natural instruments for the organization and proliferation of each movement. A diverse constellation of groups and individuals utilize these platforms to organize the actions of their political collective; ultimately trafficking in fakenews and propaganda to draw attention and motivate action in support of their movement.

Furthermore, other examples of increased political radicalization are not difficult to find if one looks towards the internet. In another instance of massive radicalization are conspiracy theories like the infamous Q-Anon theory which became a polemic issue during the Covid-19 pandemic. These conspiracy theorists frantically connected dots between all sorts of seemingly innocuous coincidences, always guided by their ultimate belief that former president Trump had been the true victor of the 2020 election, and therefore the duly elected leader of the US. These conspiracy theorists' views directly contradicted the factual narrative maintained by traditional media outlets to such an extent that they were ultimately incompatible; the mainstream media was inexorably bound to their notion of fakenews. This radically divergent ideology motivated the civil unrest and riots seen at the US Capitol in 2021, ultimately leading to a mass-scale breach of security measures and outbreaks of violence.

This newfound relationship to the digital medium which was thrust upon people due to the movement restrictions imposed in response to the Covid-19 pandemic coincides strongly with an unprecedented level of polarization and civil unrest in the US. Moreover, before the pandemic, polling data would indicate that the US was already on course for a level of ideological divergence which would certainly precede massive outbreaks of civil unrest. Furthermore, it seems that social media and the trafficking of fakenews content appears to be at the root of the most notable civil crises of late.

v) *Wearing Rose Colored Glasses: Post-Truth and the Mediation of Facts by Beliefs*

In his 2019 article, philosopher M. Camacho tracks the unsettling rise of the “post-truth era” in the US since it was first conceptualized in 1992. Here, post-truth is described as “circumstances in which objective facts are not as influential in terms of forming public opinion, as an appeal to personal emotions and beliefs” (Camacho, 2019) might be. This term seems to adequately illustrate the situation as described above. No longer do Americans appeal to basic

facts when they conduct their political discourse, instead, they depend more heavily on beliefs and emotionally charged internal principles. Importantly, we must note that the latter two are generally invulnerable to contradictory claims, they rely on something other than substantiated facts or arguments, namely faith.

Instead of factual arguments, Camacho argues that the post-truth era has made people “become accustomed to living in an environment in which the truth is discriminated against depending on our personal interests” (Camacho, 2019). If we take into account the previous two studies, we can easily observe that Camacho’s observation is substantiated. If we can interpret from section (III,ii), that the American people have deeply divergent core political values, and from section (III,iii) that a disproportionate majority of Americans do not agree on basic facts about the current state of affairs, then Camacho’s assertion might hold true. If the ideological divergence described in Pew’s 2014 study accurately indicates such a drastic divergence in political values, and most Americans would agree that their peers do not hold the same facts to be true in accordance to the 2021 study; then perhaps they truly are mediating their conception of the truth through their political beliefs, albeit inadvertently. If their political beliefs are broadly incompatible, and their understanding of basic facts are broadly incommensurate, then we should assume Comacho’s assertion on the basis that truth and facts are necessarily a shared feature of the objective world. The constitutive facts which describe the current state of affairs are common to everybody. Therefore, if the divergence on what these facts are is so great, and fall along partisan lines, we should also assume that the discrepancy is rooted instead in belief structures. From this perspective, the problem is the interpretation of these common facts, through the mediating lens of personal political ideology, thereby distorting one’s perceptions of that shared world in accordance with said beliefs. In other words, those participating in the American political dialogue appear to be approaching the discussion while wearing rose colored glasses, mediating what they see through the lens of their own beliefs.

Camacho goes on to suggest that “whoever wishes to influence public opinion should concentrate on the creation of discourse that ... place[s] an emphasis on what will satisfy the emotions and beliefs of the audience, instead of real facts” (Camacho, 2019). Camacho’s warning also tracks well against the recent history of civil unrest in the US outlined in section (III,iv). Between the massive outbreaks of violence on one hand, and the conspiracy minded thinking which has taken over swaths of the American electorate on the other; passionate belief

appears to be the *modus operandi* of political change and cooperative action in the US. Whether it be a passionate defense of minority communities with respect to an oppressive police state; or the ardent rectification of a stolen election combined with a feverish desire to defend the republic. Both instances have been characterized by their impassioned defense of what they believe to be a systemic failure, and both are wrought in emotional discourse.

This appears to be a root cause of the broadly incompatible views of the American people. Belief and emotions are not subjected to the same argumentative formulation as facts which might be asserted, defended or contradicted. Beliefs and emotions rather, are not comparable to each other, they are not easily measured, nor neatly contradicted. In so, if we take the above data at face value, there is strong reason to suggest that the American people have broadly abandoned factual argumentation as their primary mode of political discourse, instead, endorsing impassioned expressions of belief and emotion as the preferred vehicle of political change.

If we then assume that these radically divergent worldviews are motivated by differing conceptions of basic facts, a byproduct of the post-truth era the US is undergoing, we can readily see how the pervasiveness of fakenews content plays a primary role in driving this divergence. Through this interpretation, fakenews could be considered the sedimentary foundation of today's political polarization in the US. The cause of the deepest divergences, upon which each side constructs its own political narrative. Being built on wholly distinct foundations of beliefs, ones which are sustained by a constellation of fakenews content masquerading as facts: the terms constitutive of the emergent discourse are naturally becoming incommensurate.

Section (IV): A Theory of Communicative Action - Key Concepts

In his 1984 work, Habermas provides a critique of modernity's conception of reason. According to the author, this instrumental conception of reason is far too narrow to truly capture the broadness of human reasoning. Through the course his work, Habermas provides an alternative conceptualization of reason, one grounded in a neo-aristotelian approach, that of communicative action. This approach presumes that a community of reasonable actors is working in conjunction in order to achieve any number of goals, always beyond mutual understanding itself. Habermas argues that reason is, more often than not, something held

between a set of rational actors who are attempting to coordinate their actions to some degree or another. According to him, this model of reasoning avoids the greatest pitfalls of instrumental conceptions of reason, namely, their objectification of the world through the lens of means-ends thinking.

The centerpiece of his life's work, Habermas' theory of communicative action concerns what he believes is the real way that people use reason, between each other and not by themselves. Central to the exposition of his theory is the presupposition that the achievement of "understanding in language is... introduced [merely] as a mechanism for coordinating action" (Habermas, 1984a; 94) between rational actors. In so, we must assign some extrinsic purposiveness to the use of language; never is it used for the attainment of knowledge itself or other impractical types of understanding. According to Habermas, communicative action can be understood as a sort of "rational infrastructure of action oriented towards reaching [a mutual] understanding" (Habermas, 1984a;106) with regards to the realization of specific goals which require the cooperation of various rational actors in a shared world. This orientation towards interpersonal understanding and cooperative action is at the core of communicative action; that multiple actors see the need to coordinate their actions, in unison, through the vehicle of argumentation and rationality, with the ultimate goal of realizing a particular task in the world.

The remainder of this section will be dedicated to explaining concepts which will later be crucial for sketching out a Habermasian approach to the current state of affairs with respect to fakenews and the polarization of political discourse. First and foremost will be a brief explanation of some crucial elements of speech act theory inasmuch that they might pertain to Habermas' work in subsection (IV, i). Following this, subsection (IV, ii) will highlight important features concerning Habermas' theory of argumentation. Thereafter, a Habermasian conception of instrumental reason and the system will be sketched briefly in subsection (IV, iii). Finally, Habermas' conception of communicative action and lifeworlds will be explained in detail in subsection (IV, iv).

i) *Theoretical Underpinnings: Austinean Speech Act Theory*

Speech act theory is a pragmatic approach to the philosophy of language developed by J.L. Austin in his work *How to Do Things with Words*. Habermas takes this conception of

language as the primordial framework of interpersonal communication behind his conception of communicative action. Habermas however, neglects to provide an adequate explanation of speech act theory through the course of his work, therefore, this thesis will appeal directly to Austin's seminal work instead.

Austin's conception of language fundamentally rests on the conventional reasons that people use language, rather than how it might articulate some abstract conception of meaning. From this perspective, the meaning of any specific utterance can be found in a tripartite division of its constitutive components. These components are the locutionary act, the illocutionary act, and the perlocutionary force behind the use of language in real world settings. The following section will first explain the tripartite division in speech acts developed by Austin. Then, it will discuss the importance of social convention with respect to the Austinean conception of speech acts.

According to Austin, a locutionary act "is roughly equivalent to uttering a certain sentence with a certain sense and reference which is roughly equivalent to 'meaning' in the traditional sense" (Austin, 1962; 108). This refers to all of the purely linguistic features of a speech act, namely, its semantic and syntactic properties, and the medium in which it is transmitted. Furthermore, Austin holds that to perform a locutionary act is "*eo ipso* to perform an illocutionary act" (Austin, 1962; 98). Austin argues that as one performs a locutionary act, they inevitably present it with "a certain (conventional) force" (Austin, 1962; 108). This force can take the form of an 'order' or a 'request' depending on the conventional features which characterize the use of specific words, ultimately determining its meaning. These two acts come bound together in the speech act; one holding content, and the other controlling presentation. Finally, the perlocutionary force is that which "we bring about or achieve *by* saying something" (Austin, 1962; 108). These are the consequences which are incurred by the performance of a speech act, that is, outcomes like 'persuasion' or 'obedience'. Together, these three parts triangulate the true meaning of a given speech act according to Austin: its content, its presentation, and its reception.

Furthermore, Austin regulates the use of speech acts under a strict condition of conventionality. According to the author, all speech acts are situated squarely inside of social conventions. Here, Austin aims to highlight the highly contextual element of speech acts which contribute to their meaning. Specifically, that these sorts of actions must be couched in a

conventional setting; that they be performed with the intention of realizing a specific social action or convention. Without such a conventional setting “there cannot be an illocutionary act” (Austin, 1962; 118). Being fundamentally constituted by those extralinguistic features which are present in the broader social context [such as ‘ordering’ versus ‘asking’] of a speech act; the illocutionary act would be absent if the conventional setting were removed from the equation. In so, the tripartite division of meaning designed by Austin quickly falls apart; the other two sides of the triangle being incapable of wholly establishing the meaning of any speech act as we understand them.

In summation, according to Austin, in the real world we more often than not communicate with each other through speech acts. This is opposed to traditional conceptions of the philosophy of language which might claim that language possesses or refers to concepts themselves. This conception of speech acts is deeply rooted in the real world use of language; one of its principal constituents of meaning being the particular conventional context in which speech is articulated in the Austinean system. The meaning, therefore, of language used by people on an everyday basis is triangulated between: the linguistic vehicle used, the context they are presented in, and the consequences which emanate from the speech act. Austin claims that only through this tripartite pragmatist approach, may the true meaning of someone’s expressions be understood.

ii) *Habermas’ Argumentation Theory:*

The next step crucial to understanding Habermas’ work is his conception of argumentative speech acts and argumentation in general. According to Habermas, “arguments are the means by which intersubjective recognition of a proponent's hypothetically raised validity claim can be brought about and opinion thereby transformed into knowledge” (Habermas, 1984a; 25). In other words, arguments are the tool by which we elevate mere opinions to the category of knowledge via the judgment of other rational actors. Furthermore, Habermas stipulates that “we can distinguish three aspects of argumentative speech” (Habermas, 1984a; 25): its procedural elements, its normative regulation, and the general aims of this type of speech.

According to Habermas, argumentation must be conceived of as a diffuse process of “reflective continuation, with different means, of action oriented towards reaching and

understanding” (Habermas, 1984a; 25). Here, Habermas indicates that the argumentative process is one which multiple rational actors reflect on the subject at hand, each independently and via their own means; ultimately bound together by their mutual intention of reaching a common understanding. Furthermore, this process is conditioned by the fact that all participating actors “presuppose in general that the structure of their communication... excludes all force... except the force of a better argument” (Habermas, 1984a; 25). This implies that all participants of an argument purge any self-interest which could arise from either internal or external sources. These sources might be a closely held principle in the case of the former, or perhaps related to some sort of objective interest in the case of the latter. In summation, the only guiding principle in an argument should be the purely “cooperative search for the truth” (Habermas, 1984a; 25) between its participants.

Furthermore, Habermas conceives of the argumentative space as one which is normatively regulated in the form of “a cooperative division of labor between proponents and opponents” (Habermas, 1984a; 25) which mutually aim to verify the validity of the claims made by participants. In so, no third party arbiter has a superior argumentative force invested in their assertions. This notion of participant parity is of crucial significance to the effective evaluation of arguments. Furthermore, Habermas specifies that the argument must be normatively regulated in a way which guarantees three outcomes. Firstly, that the participants successfully “thematize a problematic validity claim” (Habermas, 1984a; 25). This implies that the argument’s participants be able to effectively flesh out the problematic elements of a validity claim, and further, establish whether or not they stand as a proponent or an opponent to the claim. Secondly, Habermas indicates that participants must be “relieved of the pressure of... experience” and operate from a purely “hypothetical attitude” (Habermas, 1984a; 25) with respect to the arguments given by others. This means that all participants must purge their judgment of prior prejudice which they might hold due to their life experiences: that my musophobia not influence our argument over the ethical implications of animal research, for example. Habermas’ final condition is that arguments be tested “with reasons, and only with reasons, whether or not the claim defended by the proponents rightfully stands or not” (Habermas, 1984a; 25). Here, Habermas highlights the fact that reasons can be the only measure of an argument’s force, and that moreover, that it is the duty of the claim’s opponents to test this matter. Habermas does not provide a strict set of normative principles which reach these three goals, instead, entrusting those participating to self-regulate

themselves towards these specific outcomes. In so, he allows for a diffuse method to emerge naturally between those arguing, without imposing any rigid structure of his own, and guided only by the force of arguments themselves.

The final crucial element to argumentative speech is its intention to produce verifiable validity claims which are to be evaluated between participants. Here, Habermas stipulates that argumentative speech always aims “to produce cogent arguments that are convincing in virtue of their intrinsic properties with which validity claims can be redeemed or rejected” (Habermas, 1984a; 25). This is the preordained goal of argumentative speech itself: the development and testing of coherent and defensible claims which are understood to be true; and moreover, vulnerable to the verification of their veracity by those participating in the argument. In so, over time, participants might find their way closer towards the truth, or at the very least a mutual understanding of the same. Additionally, Habermas identifies three conditions which must be met if a claim is truly vulnerable to having its validity checked by an argument’s participants. Firstly, “that the statement made is true” (Habermas, 1984a; 99), implying that the claim be indicative of a real belief held by the claimant with respect to the ongoing state of affairs. Secondly, validity claims are articulated via a “speech act which is right with respect to the existing normative context” (Habermas, 1984a; 99), namely, those speech acts of an argumentative character. Finally, these sorts of claims must be made in a way in which “the manifest intention of the speaker is meant as it is expressed” (Habermas, 1984a; 25), in other words, that all participants be expressing their beliefs wholeheartedly. If an argument can be subject to these three dimensions of critique, then it may be considered vulnerable to having its validity tested by those participating in the argument. This goal of reaching mutual understanding conditions any articulation of argumentative speech; it is always oriented towards the production of arguments which are both convincing to a plurality of rational actors, and subject to validity claims.

Habermas conceives of argumentative speech as a special sort of speech act which occurs within the social convention of ‘argumentation’. Translated into Austinean terms, the illocutionary qualities of speech acts might be considered to be argumentative, if and only if, they comply with the conventional requirements set in place by Habermas. Specifically, a speech act is considered argumentative, if it is contextualized within a specific argumentative procedure, if it is normatively regulated by one’s co-participants, and is aimed at the production of

convincing claims which are subject to the verification of their veracity. By combining an Austinean conception of speech act theory, with his own theory of argumentation, Habermas is able to effectively qualify which speech acts can be considered purely argumentative.

iii) Instrumental Reason & System

In order to adequately comprehend Habermas' conception of communicative action, its contrary, instrumental reason must first be addressed briefly. As opposed to the model of communicative action, instrumental reason operates from the standpoint of a solitary thinker, one who utilizes reason to mold the world around their interests. Habermas admits that the "avoid[ing] errors in regard to facts and means-ends relations" (Habermas, 1984a; 15) is certainly a fundamental feature of traditional conceptions of rationality. However he also fixates on the fact that this is an overly narrow conception of rationality, that "there are obviously other types of expressions for which we can have good reasons" (Habermas, 1984a; 15). For the purposes of this thesis only the mode in which this type of rationality treats the world, and how this then leads to the systemic reification of instrumental reason will be attended to.

In simplest terms, instrumental reason is that type of rationality which relates to the world on the grounds of subject-object relations, wherein everything is understood through the lens of means-ends thinking, thereby implying the objectification of the world over time. Habermas argues that the tradition of instrumental reason merely aims to "reduce practical reason to a competence for the purposive-rational choice of means" (Habermas, 1984b; 213) in relation to meeting specific ends. From this perspective, instrumental reason is that rational infrastructure which leads us to the selection of 'best practices' in relation to specific goals. Through such a lens the subject takes on a monological relationship to the world, considering all those things outside of their own subjectivity, objects, which are merely means by which to achieve particular interests. In this instrumental space, Habermas warns that "interactions are no longer coordinated via norms and values, or via processes of reaching understanding, but via the medium of exchange value" (Habermas, 1984b; 336). This value takes on an objective character under instrumental thinking; oftentimes money, power, or resources in a social setting. This means that individuals are incentivized to comport themselves in accordance to what means might achieve the greatest ends, objectively, through the model of instrumental action. This derivation from an

intersubjective approach lends itself to the wholesale objectification of the world, up to and including other people.

Habermas understands that furthermore, instrumental reason becomes broadly reified through society, eventually engendering a tangible infrastructure which stands in opposition to one's worldview (or lifeworld in Habermasian terms): the system. According to Habermas, through instrumental reason, individuals “orient themselves to ‘values’ in a purposive-rational manner, as if ... [values] were objects in a second nature, they adopt an objectivating attitude to each other and themselves, and transform social and intrapsychic relations into instrumental relations” (Habermas, 1984b; 336). According to Habermas, this type of means-ends thinking is infectious, it continually leads to further objectification, moreover, the transformation of social relations into instrumental ones. This process of transformation aims to capture the entirety of the lifeworld, slowly consuming its communicatively generated elements. With this natural impetus, the system carries out “a methodological objectification of the lifeworld and [a] justified shift in perspective connected with this objectification— a shift from the perspective of a participant to that of an observer” (Habermas, 1984b; 374) in terms of action coordination. This shift away from intersubjectivity bars an individual from effectively participating in the world, instead reducing him to a mere observer, one who is sitting in front of a silver screen as lurid images of a bustling society breeze by him. From this perspective, the system is far more rigid than the lifeworld: characterized by objective, and thereby insurmountable institutions, whose values are absolute and actions merely strategic.

Habermas’ conception of instrumental reason, and thereafter the system which is engendered by its reification, bears a dichotomous relationship to the lifeworld. From this conception, the totality of the world is characterized by the lifeworld of communicative social relations on one hand, and the objectifying institutions which are wrought in instrumental rationality on the other. Born from totally opposed ingredients, Habermas criticizes the utterly toxic characteristics of instrumental reason and its kin. The strategic utilization of others for the achievement of one’s personal interests runs in stark contrast to foundational principles of the traditional western conception of deontological ethics, namely the Kantian categorical imperative which demands we treat people with intrinsic value, as ends themselves, and not means. Given this moralistic fact, Habermas warns that the objectifying effects of the system represent an ever

encroaching force upon the intersubjective lifeworld, slowly but surely consuming any chance of communicative action in favor of an unyielding paradigm of instrumental reason.

iv) Communicative Action & Lifeworld

Communicative action constitutes Habermas' response to the conceptualization of instrumental reason provided by modernity. Describing instrumental reason as a barebones conceptualization of rationality, one which is deeply flawed due to its fixation on mean-ends thinking and objectification. Habermas aims to provoke a paradigm shift in relation to our understanding of reason, towards a model of communicative action rather than instrumental reason. The former model presupposes the existence of a community of rational actors with shared interests, whereas the latter operates under a model of the solitary thinker. The purpose of such an activity is twofold, it addresses both the teleological aspects of realizing one's goals, and the communicative aspect of interpreting the world and reaching mutual agreement (Habermas, 1984b; 126). To do so, Habermas binds the two through rationality; the coordinative process of the former with the argumentative approach of the latter. According to Habermas one cannot survive without the other: cooperative action is doomed to fail without prior argumentation, and communication rendered meaningless without a concrete orientation towards the world. In so, the model of communicative action presents itself as the only way to orient an agent towards the real world, through reason, without relying on a purely instrumental mode of thinking. Habermas argues that this is the way that people really use rationality in their everyday lives. People most often use reason, through the vehicle of argumentation and discourse, and oriented towards some sort of external purposiveness, with the stated intent of reaching mutual understanding and thereby successfully realizing coordinating their actions in the real world. To achieve these goals, communicative action relies on the exercise of communicative rationality and consensual agreement in order to realize cooperative activities. Eventually, repeated use of this model leads to the formation and evolution of the Habermasian lifeworld as its backdrop.

Communicative rationality can be taken to be an intersubjective approach to the problem of reason; primarily focused on the achievement of mutual understanding through the exercise of effective argumentation. According to Habermas, communicative rationality is a process by "which different participants overcome their merely subjective views and, owing to the mutuality

of rationally motivated conviction, assure themselves of both the unity of the objective world and the intersubjectivity of their lifeworld” (Habermas, 1984a; 10) through the vehicle of argumentation. In so, those participating in an exercise in communicative rationality are able to determine objective facts about their world with the assistance of their rationally minded peers, via a model of intersubjective observation and discourse. Here, participants are able to affirm or reject validity claims made by others on the basis of an effective argumentative model as outlined in section (III, ii). In so, the group can establish factual truths about either the objective or intersubjective worlds in a consensuated fashion, slowly forming a consistent background of mutual understanding, upon which later judgements will be contrasted.

Habermas’ conception of the lifeworld is constituted by this all-encompassing background of mutual understanding. The lifeworld is conceived of as that naturalistic environment in which ordinary social procedures and linguistic communication bear their true weight. Highly contextualized by the shifting sands of human culture and tradition; expressions gain meaning based on the particular socio-conventional context that they are used in. According to the author, “the lifeworld is the intuitively present, in this sense familiar ... and at the same time [a] vast and incalculable web of presuppositions that have to be satisfied if an actual utterance is to be at all meaningful, that is, valid or invalid” (Habermas, 1984b; 131). Habermas describes the lifeworld as the familiar, yet imprecise and diffuse network of socio-cultural traditions which imbue our speech with meaning. Not only does this indicate those intangible features which give simple terms such as ‘police’, ‘state’, or ‘president’ their definition, but also engenders the specific connotation related to their actual use. From this perspective we can easily observe how ‘police’ might mean ‘oppressive force’ for certain individuals, meanwhile indicating ‘law enforcement agency’ for others. Furthermore, the lifeworld serves as “the unquestioned ground for everything given in ... [one’s] experience, and the unquestionable frame in which all the problems ... [one has] to deal with are located” (Habermas, 1984b; 131). One’s lifeworld is that apparently objective frame of reference from which all judgements are interpreted; made to be invulnerable from critical attitudes. Despite this unflinching certainty in the monologic dominance of one’s lifeworld over their judgements, this does not impede efforts in communicative action due to its reliance on intersubjectivity. Indeed, Habermas affirms that “the lifeworld owes this certainty to a social a priori built into the intersubjectivity of mutual understanding” (Habermas, 1984b; 131). This debt is due to the necessarily intersubjective

evolution of culture and language, neither can be progressed monologically. No culture is expressed for the sake of expression, nor language spoken for the sake of oneself. Rather, due to the dialogical nature of cultural development these presuppositions which are seemingly objective to the individual, are actually vulnerable to the intersubjective the interrogation and revision of a rational community. In so, these objective-feeling presuppositions are ultimately rendered malleable by the mediation of other rational actors through the vehicle of discourse and argumentation. In so, the lifeworld provides the quasi-rigid backbone to one's ability to make judgements about the world, fundamentally determining their conception of truth and meaning through the lens of a priori intersubjective agreements. In summation, the lifeworld is that objective-feeling a priori shared framework which has been constructed by repeated acts of communicative rationality; constituted by those intersubjective assumptions which individuals use to mediate their judgements.

According to Habermas, communicative action can be understood as the coordination of activities between multiple rational actors, through an argumentative model which is rooted in the intersubjective approach of communicative rationality. At its core, communicative action is the use of communicative rationality as the fundamental driver of cooperative action. In so, communicative action is capable of utilizing reason, while respecting the social norms and realities which are constitutive of social activities, namely, the coexistence of rational actors, as opposed the solitary thinking inherent to instrumental conceptions of reason. Furthermore, successful attempts at communicative rationality slowly engender a vivid lifeworld; one constituted by the history of mutual agreements ordained by the prior arguments of a rational community.

Section V: *A Habermasian Approach to the Problem at Hand*

Now that both the problems of fakenews propaganda and ideological polarization in the contemporary US have been addressed alongside a few key concepts of Habermas' work: a Habermasian approach to the issues at hand can be developed. The remainder of this section will be dedicated to redescribing the problems listed in sections (II) and (III) through the lens of communicative action in subsection (V, i). Then, will continue on to conceive of possible

solutions which could be articulated via a specifically Habermasean approach in subsection (V, ii).

i) Divergent Lifeworlds & the Abandonment of Communicative Rationality

If approached from a Habermasean perspective, the fundamental problem which is currently plaguing the political discourse of the US appears to be rooted in how ‘truth’ and ‘facts’ are used by those participating. It would appear that due to the emergence of fakenews, political discourse had become tainted, the argumentative procedure behind communicative rationality broken, resulting in a degradation of the value of validity itself within the discussion. In so, what were once moderately divergent political values in the 1990s, have become incompatible lifeworlds in the 2020s. I argue that the breakdown of standard argumentative procedure given the newfound instrumental value of the fakenews, has fueled the rapid ideological divergence which had already been developing between competing political factions in the US; ultimately rendering the presupposed structure of consensuated agreements (lifeworlds) between participants wholly incompatible, and thereby has provoked a meltdown of the country’s political discourse.

This breakdown of communicative rationality is rooted in the fact that partisan beliefs have taken on the role of previously consensuated facts in the argumentative process described in section (IV, ii). This failure is articulated through the assertion of beliefs as facts, such as: ‘the police *are* an oppressive force’ or ‘the election *was* stolen’. Such certainty serves to disarm a claimant’s opponents, impeding their ability to check the veracity of a given validity claim, precluding the possibility of any fruitful agreement with respect to it. There are three argumentative failures related to this formulation with respect to communicative rationality. Firstly, that in the former scenario, the claim is not a validity claim, its veracity is not vulnerable to interrogation by participants. Secondly, these claims do not presume that a mutual understanding can be reached between participants, instead, it is a statement of fact irrespective of contrarian beliefs. Finally, these claims are not articulated through a hypothetical attitude, instead, they operate as on the basis of objective knowledge: infallible certainties. These failures in the discursive protocol are rooted in the misinterpretation of beliefs, as facts within argumentative procedure. In so, the process of communicative action is broken at its core; the

normative system which regulates argumentation is hollow. From this perspective, we should consider fakenews content the vehicle by which systemic influence is smuggled into the lifeworld: a parasite at the core of communicative action.

From this point, it is important to address the divisive nature of fakenews content; having fueled a substantial radicalization on both sides of the aisle. This polarization has led to the emergence of two ideologically incompatible communities as described in sections (III, i; III, ii); furthermore, groups who disagree on basic facts as seen in section (III, iii). From this standpoint, we would do well to envision fakenews as the tool by which systemic influences might artificially divide a shared lifeworld, and furthermore, to actively prevent its reunification. We can observe these divergent lifeworlds in the newfound propensity of online echo chambers described in section (II, ii). These echo chambers then operate as a rubber stamp, objectively validating the veracity of fakenews content *en masse*, from an apparently intersubjective authority. In so, instrumental reason is operating within the core of one's lifeworld, continually calcifying intersubjective understanding into objective fact, rendering the possibility for discussion inert. Fakenews content serves not only to divide a rational community at its core; but also to slowly but surely solidify those differences through the reification of belief as fact within one's lifeworld.

Concurrently with Comacho's illustration of the post-truth era described in section (III, v), I argue that the value of truth has deteriorated in the contemporary political discourse of the US, more so since the emergence of social media platforms which provide safe harbor to fakenews content. Due to flagrant misuse of the argumentative procedure which is facilitated through the spread of fakenews on social media networks; truth and belief are oftentimes confused. In so, Americans see themselves forced to abandon discourse as the primary mode of realizing political change given its inability to confirm the veracity of what are presented as validity claims to participants. Instead, resulting in a positive feedback loop wherein fakenews inserts itself as fact, provoking the claim's rejection by one's opponents, which then in turn reinforces the veracity of the initial fakenews given the psychological tendency towards confirmation bias. This feedback loop slowly jumbles what is and is not considered truth, until two incommensurate ideological factions are sifted out; each believing in a system of objective facts which is incompatible with that of their opponents.

This total failure of communicative rationality to realize its central task (the reaching of mutual understanding through proper argumentation), which is due to the introduction of fakenews into our information bearing, has resulted in the slow collapse of political discourse in the US. Instead, confronted by the total breakdown of communicative action, Americans have appealed to impassioned ruptures of civil peace as the primary vehicle through which to advance their political aims as seen in section (III, iv). Incidents like the George Floyd riots which aimed to tackle police reform; or the Capitol riots which aimed to enact an electoral change. However, rather than approach the dilemma from the perspective of communicative rationality, both sides embraced an emblazoned rhetoric wrought in a system of authoritative facts; which has served to instrumentalize the public in an attempt to incite political change. The failure of communicative rationality has not altered the teleological aspects inherent to the pursuit of political change, but rather, has pushed them into the domain of instrumental action.

Approaching the problem of fakenews and political polarization from the standpoint of Habermasian conception of communicative action lays out various deeply problematic issues within the context of ongoing political discourse. More than merely memes or misinformation; fakenews is a parasite which has attached itself to the core of communicative action in American society. In simple terms, the way that Americans use fakenews in their political discourse is conventionally flawed, they instrumentally misappropriate the value of different elements of argumentative procedure: they smuggle belief inside of argumentative speech acts. This has nullified the argumentative procedure in the context of political discourse in the US, rendering it incapable of realizing its truth validating purpose. It has rendered the political discourse wholly inert in the country, pushing Americans towards sporadic outbreaks of civil unrest in lieu of discursive pathways towards political change. More so, the proliferation of fakenews in the American political discourse has engendered two ideologically incompatible factions within the electorate, moreover, keeping them divided at all costs. Under such conditions intersubjective political discourse is not a viable path towards political change; only systemic influences and objectifying instruments will successfully motivate reform.

ii) *A Habermasian Solution to a Digital Age Problem*

Given the description of affairs provided in the previous subsection, the question remains, how might this situation be repaired? If we are to continue following the work of Habermas, the solution may be found in the revitalization of communicative action between those people who are participating in the American political discourse. That they see political discourse as the most effective path towards political change, as opposed to instrumental action. To do so, a system which can provide grounds ripe for communicative rationality must be developed, one which can adequately account for the truth-deteriorating effects of fakenews content. I argue that, due to the pervasiveness of fakenews online, a digital platform would be the best means of fostering said environment: an app.

If formulated appropriately, such a platform could do away with the most toxic and truth-deteriorating features of the current political discourse which were noted previously; while simultaneously eroding the ideological barriers reproduced by the fakenews system. In order to achieve such a goal, certain features which are stereotypical of the current model of social media should be addressed, specifically, those features which impede the exercise of communicative rationality. As discussed above, the root of this problem appears to be found in the failure to produce a “cooperative search for the truth” (Habermas, 1984a; 25), not the coordination of action towards specific ends.

In order to construct a platform which could facilitate the exercise of communicative rationality, several key elements must be addressed, namely, the mode in which information is presented on the network, the normative regulation of discourse, and measures which increase the social accountability of its users. The following section will discuss specific features which could be implemented in a social platform to enhance user’s ability to exercise communicative rationality effectively in the terms of political discourse.

Section VI: *Speaking to Each Other, Rather than @eachother*

In order to facilitate the use of communicative rationality in online political discourse, several key issues must be addressed. These issues concern the corruption of the argumentative procedure as discussed in subsection (V, i). The following section will discuss three crucial features which would greatly enhance a user’s ability to cultivate fertile grounds for argumentation, and therefore the assertion and critique of validity claims. To do so, iconic

elements of social media platforms must be addressed, such as: forum posting, user comments, and user anonymization. These features, characteristic of sites like Twitter, Facebook, and Instagram allow users to evade argumentative convention via an instrumental approach to social communication, and thereby traffick in those validity-invulnerable claims, fakenews. By substituting these features with one's which should more effectively facilitate communicative rationality, a social media platform could be designed in which users speak to each other discursively, rather than merely slinging what appear to be facts at each other from their respective towers.

Such an application would necessarily take the form of a group messaging platform; fostering a space where social peers or colleagues can create small groups and discuss content which they find to be of relevance to the shared interests of its members. These small groups would be much more pliable in terms of cultivating grounds ripe for communicative rationality, facilitating the exercise of those truth-seeking elements of argumentation. In so, users could work together with their peers to evaluate the veracity of claims made online. Such an app would naturally capture a user's attention due to the a priori shared interests of its members who have consensually joined the group under the guise of curiosity in given a certain topic. In so, users would be incentivized to share content directly with the small group, inviting argumentation and discussion with respect to the validity of the content. Slowly over time, groups would develop a repertoire of valid and invalid claims which might be reviewed or cited by individuals in other discussions on or offline. By shifting the focus of the platform's primary system for disseminating information away from massified 'posting' and towards the creation and small interest groups constituted by real world relationships, fertile grounds could be developed for the exercise of communicative rationality.

This section does not aim to wholly design a functioning social media network, instead, it aims to provide constructive criticism of the traditional features which can be considered nearly universal amongst today's platforms. In so, offering a perspective into how they could be redesigned in a way which incentivizes communicative rationality, rather than merely impeding it.

i) *Replacing Forum Posting: Cultivating an Argumentative Space*

Forum posting can be understood as one's 'timeline' on Facebook or 'feed' on Instagram and Twitter. It is that infinitely long list of posts which users scroll through aimlessly for hours on end. Forum posting is the primary mode of presenting information to users on these platforms: content is algorithmically selected and slowly trickled to the user as they scroll downwards. Furthermore, this model can be seen as an objective mode of presenting content to its users; presented monologically to an audience who are underrepresented with respect to their argumentative status. The poster naturally takes a predominant position within the discussion, their opinions and beliefs being commented on primarily. In so, there is a seemingly bilateral discussion between commenters and posters, however, there is no real intersubjective discussion of a rational community. Rather, each comment takes the form of an isolated critique which can be acknowledged or ignored by the poster, rather than responded to. Moreover, due to the vastly disproportionate number of commenters versus posters on viral posts, the latter is naturally inclined to ignore a great number of these comments, being logistically impossible to address them all. This dynamic fundamentally undermines the parity of between claimant and opponents which is inherent to a Habermasian conception of argumentative theory. Even graphically, the 'post' is disproportionately larger than the comments which refer to it no matter the argumentative validity of a counterclaim, cultivating a cloud of invulnerability around the posted content. This aura of invulnerability severely debilitates any efforts in communicative rationality which might work to unravel the truth of the matter.

To substitute this, another dynamic must be cultivated, one grounded on intersubjectivity as opposed to objectivity. This goal might be advanced most effectively via the wholesale substitution of forum posting in favor of a framework which disseminates content through a multiplicity of small groups, in so, reining in the objectifying dynamic of a 'post'. Similarly to the way an instant messaging platform functions, users would submit content to a small group of peers into a space the likes of a chatroom, wherein posts and comments take up the same conventional space. Through this method, the objective value of a post is deteriorated, and made vulnerable to the intersubjective critique of peers. Furthermore, a platform which is centered around the formation of small groups between social peers would highlight the vulnerability that each claim has to a discussion's co-participants; no longer is one forced to assume that the poster is either highly knowledgeable or utterly moronic with respect to the content they have uploaded. In so, content is made malleable once again to the forces of intersubjective elucidation.

ii) *Replacing Comments: Normative Regulation of Argumentative Speech Acts*

Another glaring issue with regards to the current formulation of social media platforms is that they fail to foster an environment where speech acts are regulated by a peer group. As it currently stands, there are two primary methods for a community to regulate the validity of a post; either via some sort of unary voting system ('likes' for example) or via comments. Neither the monofocal metric offered by these platforms, nor the barebones commenting features serve to adequately regulate the speech acts used online. These meek tools are easily ignored in the case of the latter; or simply serve to regulate a content's exposure to the community in the case of the former. These two methods fail to establish a normatively regulated environment: the consequences of rule violation are too lenient to proactively cultivate an adequate argumentative space. Comments should serve as critique, the rejection of a claim by a member of an online discussion should prompt argumentation, instead, they merely pile up like emails in one's spam folder. Both of these features fail to effectively regulate users' speech acts in accordance with argumentative convention.

A direct consequence of replacing forum posting is subsequently the removal of comments as they are currently conceived, as seemingly critical parasites which are attached to posts. In virtue of the substitution of forum posting, argumentative convention is more easily regulated within these small groups of colleagues as well. A violation of argumentative procedure, such as the misuse of beliefs as facts, can be met with more impactful reprisal by one's opponents if there are so few. Decreased peer group size leads to each member holding a proportionately larger share of the collective authority and therefore influence; thereby increasing the dialogical value of their comments. Moreover, repeated violations of normative procedure can lead to social consequences that go beyond punitive measures such as temporary censure or removal from the group, instead; real world relationships are at stake. In so, the argument is oriented inexorably towards the real world, its participants bound to each other via socially connected peer groups. In so, self-regulation of the argumentative space is made possible through the paradigm of small groups as opposed to forum posts and their respective comments.

iii) *Dissolving Anonymity: Engendering Accountability between Peers*

One of the indissoluble pillars of internet communication is anonymity; it is nearly impossible to identify someone who refuses to be identified online. Platforms like Twitter are infamous for anonymous posters, bots, and throwaway accounts which are fabricated for the purpose of violating community standards. This anonymity provided by the digital medium is utilized by users to mask themselves online, rendering them immune to the ordinary social consequences which might result from misbehaving so gravely. In so, they articulate claims objectively, unlinked from any particular person who might have acted in such a way. This way, merely provocative or sensational claims are incentivized; those which create controversy and attract attention. Users are therefore prone to express those claims which they would not otherwise express in an offline argumentative situation, one characterized by the presence of socially interconnected peers. Hidden behind the guise of anonymity, users are not accountable to each other, rather, they are ungrounded from the social convention in which argumentation is borne.

The solution to the anonymity problem relies on the real world interconnectedness of users on the platform. In order to pursue this goal, users should be connected in offline environments (like they already are across networks like Facebook and Instagram), so that one perceives the social consequences of their [mismatching] speech acts. In so, a platform which incentivizes communicative rationality would not aim to connect previously disconnected people, but rather enhance the connection between preexisting social peers. This newfound connectedness is crucial to the wholehearted expression of beliefs which is a prerequisite to successful argumentation. In so, users would be incentivized to voice their opinions in a way which is appropriately mediated by the intersubjective perspective of the group, as opposed to the seemingly objective nature of an anonymous post. Furthermore, this a priori connectedness foments user's ability to orient themselves towards mutual understanding, being that they already share a certain constellation of social values by being in each other's social circles offline.

iv) *Speaking to Each Other Rather than @eachother*

By dismantling the central pillar behind the current design of social media platforms, forum posting, great strides might be made in the name of communicative rationality. This centralized method of disseminating information is totally debilitating to the exercise of communicative rationality. The system incentivizes the objectification of discourse while disincentivizing the fruitful participation of its audience. This system undermines the parity of participants in a debate, leading to an unregulated quagmire of dubious argumentative strategies. Furthermore, this system fails to provide a space in which users can wholeheartedly express themselves, instead, favoring the dissemination of only the most virulent claims. In so, they lead to the creation and expansion of echo chambers, which bear a polarizing influence towards real world discourse.

On the other hand, by reformulating the way information is shared on social media platforms, communicative rationality, more than just being provided a space might be incentivized. By dismantling the most toxic and objectifying elements of social media discourse, perhaps these platforms might produce fertile grounds for people to exercise communicative rationality, and thereby realize communicative action in the real world. By eliminating the objective presentation of content in the massified mode of forum posting, and replacing it with an intersubjective model oriented towards the “cooperative search for the truth” (Habermas, 1984a; 25), progress might be made in the moderation of political discourse.

Far from a comprehensive analysis of social media platforms, this evaluation of fundamental features aims to highlight concrete problems which impede the exercise of communicative rationality between users. In so, advancements could be made in the formulation of a platform which operates under the premise of communicative action, rather than one which is merely used to instrumentalize the masses towards political action. By cultivating this rich argumentative space, users might have an effective tool by which to distinguish truth from fakenews more effectively, thereby stoking the moderation instead of polarization of their political discourse.

Section VII: Risks Related to an Intersubjective Formulation of Social Media Networks

Despite its focus on intersubjective thinking as opposed to instrumental reason, this approach to political moderation is not without risk. When we consider the implementation of a

social media platform with the aim of realizing political moderation and designed around a concept of communicative rationality, two primary risks emerge: its backfire potential, and its failure potential.

The greatest risk related to the implementation of such a system is its potential to backfire; that it might provoke further polarization. It would be easy to argue that the formulation of the social media platform proposed by this thesis simply implies the mass production of echo chambers. That the strategy of grouping users based on their preexisting relationships and interests lie dangerously close to our definition of echo chambers provided in subsection (II, ii). While there is a striking similarity in the structure of the two groups; echo chambers do not rely on dialogical modes of truth-seeking, but rather monological presentation of previously validated content. The express purpose of an app grounded on communicative rationality would be the validation of claims; not merely their dissemination as is the case for echo chambers. Furthermore, this risk can be generally discounted given the existence of platforms which already serve as a better medium for mass-scale dissemination of content, such as Twitter or Facebook. In so, political radicals who aim to provoke polarization through the instrumental use of echo chambers, will rely on other platforms which are more specialized in that purpose; as opposed to the proposed platform, which aims to limit the dissemination of content to colleagues and relatives.

Another risk to such a system is that it fail due to its overly heavy reliance on continued user interest to effectively realize its goals of moderating political discourse. This risk is highly legitimate, pointing at a grave problem of any solution grounded in individual virtue: will people really use it? After all, some might claim that thoughtful discussion of nuanced topics is not something ordinary people are generally interested in. On the contrary, I argue that such an assumption is untrue, and that rather, large swaths of people would be highly interested in such discussions. This claim can be evidenced by the rise of podcast content: with shows like The Joe Rogan Experience who's total duration is hours long, rising to prominence in the late 2010s, long-form content has become resoundingly popular. Shows like Rogan's are viewed by millions of people and are understood to be highly influential, addressing a wide variety of topics through a model which already resembles that of communicative rationality. Podcasts generally include multiple participants conversing over a number of topics over the course of an hour or more; who's stated interest is the pursuit of fruitful conversation and mutual understanding.

Furthermore, inasmuch lack of interest is concerned, one could easily invoke the principle of non-maleficence. Even if such an application could not produce a meaningful shift towards moderation in political discourse, its failure does not imply harm to any parties. Rather, the argumentative weight should be carried by the opposing claim: what harm is produced by a system which incentives communicative rationality if it does not enjoy widespread use? The natural response would be none; if such a system is not broadly adopted by a rational community, then it bears no weight onto the state of political discourse.

These objections towards the implementation of a social media platform which is designed in the light of communicative action are not without refute. Despite indicating what could be considered debilitating failures with respect to its fundamental purpose, such criticisms can be adequately addressed for the reasons given above. Such a platform would not be likely to backfire, being that other platforms already exist which offer more favorable conditions from which to incite political polarization. Moreover, failure to produce a more moderate political discourse would not imply any more harm than what is already incurred by the system at large.

Section VIII: Concluding Remarks

The work of J. Habermas offers a novel approach to the problems of fakenews, truth, social media, and political discourse in the 21st century. From this perspective, we are led to believe that the way people communicate with each other through this digital medium fundamentally fails to emulate the ways in which humans exercise intersubjective discourse in ordinary social settings. That these social networks fail to cultivate an argumentative space in which the veracity of claims might be assessed by a cohort of co-participants: deteriorating the value of truth itself in the face of belief. This failure to regulate the normative structure of discourse online has contributed to the wholesale breakdown of communicative action as an effective mode of realizing political change. Instead, rational communities, entire nations, have been led to instrumental action as their preferred vehicle with which to realize their political goals. This dynamic is not insoluble however, I argue that there is a way to purge these networks of their instrumental features, a platform grounded on communicative rationality. Such an application could offer a competing platform which serves to ameliorate the political polarization introduced by fakenews into traditional social media networks. A social media platform

grounded in a Habermasian approach to communicative action, one which is focused on fixing the conventional problems at the core of how people articulate their arguments online, appears to be the most elegant solution to the problem of fakenews and polarization in contemporary political discourse, addressing the issue from the inside.

Far from comprehensive, this thesis begs for more research into the matter. The topic of social media platforms and political radicalization are subjects which are deeply in need of further study. These issues are oftentimes hard to relate comprehensively for a number of reasons. Firstly, is the drastic change in society's relationship to digital media since the Covid-19 pandemic. This thesis being published so closely to the pandemic, very few studies are able to relate the two topics in a meaningful way by utilizing data gathered from these platforms. Secondly, these topics are oftentimes conceptually difficult to relate through traditional data analysis. Studies which focus on big data analysis are oftentimes confronted with the difficulty of translating objective data points such as followers or likes into concrete shifts in political attitudes. On the other hand, ingenuitive studies designed in the fashion of Hong & Kim's work, detailed in subsection (III, i), provide rich grounds for interpretation due to their reliance on creative measurements which qualify political polarization via practical metrics such as legislative activity. Furthermore, there are few studies which take such a radical approach to the notion of truth and its deterioration as described by Comacho's article which was illustrated in subsection (III, iv). Comacho's illustration of the post-truth era and its conceptual evolution since the 1990s merits further investigation inasmuch we might consider it a material change in how the notion of truth is used broadly across society. Finally, there is an important lack of research in how speech acts are used across social media. Being focused on the conventional purpose of language, speech act theory appears to be a good avenue through which to approach the problem of fakenews: focused primarily on the force with which language is presented, and the effects it holds on its audience. Infrequently is speech act theory used to assess problems in digital communications; meanwhile it would appear as if digital speech acts are, conventionally, distinct to ordinary speech acts.

Communication online is conventionally distinct to the ways in which ordinary people articulate discourse in their ordinary social interactions. I argue that social media platforms are constructed in a way that inadvertently facilitates the propagation of fakenews content due to specific features which inhibit the creation of truly argumentative spaces, those ripe for

communicative rationality. This content bears grave consequences for the ongoing political discourse of the US, artificially separating the electorate into ideologically incompatible factions. In virtue of this separation, communicative action has been abandoned as a fruitful model for coordinating political change, instead, instrumental action has become the preferred vehicle of advancing one's goals. Despite these facts; I argue that a social media platform could be designed in the light of communicative action, one which aims to rectify those features which impede the exercise of communicative rationality. Such a platform would need to replace fundamental features of traditional social media networks with an intersubjectively minded substitute. In so, a social media platform could be designed which aims to reconcile the rapidly divergent worldviews of those factions which realize the political discourse of the US via a model of Habermasian communicative action.

References

- Al-Rawi, A. (2021). Political Memes and Fake News Discourses on Instagram. *Cogitatio*, 9(1), 1-15. DOI: 10.17645/mac.v9i1.3533
- Austin, J. L. (1962). *How to do Things with Words*. Oxford University Press.
- Barberá, P. (2020). Social Media, Echo Chambers, and Political Polarization. In J. A. Tucker & N. Persily (Eds.), *Social Media and Democracy: The State of the Field, Prospects for Reform* (pp. 34-55). Cambridge University Press. <https://doi.org/10.1017/9781108890960>
- Connaughton, A. (2021, October 13). *People in US see stronger conflict in society than in other countries*. Pew Research Center. Retrieved June 16, 2022, from <https://www.pewresearch.org/fact-tank/2021/10/13/americans-see-stronger-societal-conflicts-than-people-in-other-advanced-economies/>
- Habermas, J. (1984). *The Theory of Communicative Action: Reason and the rationalization of Society* (T. McCarthy, Trans.; Vol. I). Polity.
- Habermas, J. (1984). *The Theory of Communicative Action: The Lifeworld and System: A Critique of Functionalist Reason* (T. McCarthy, Trans.; Vol. II). Beacon Press.
- Hong, S., & Kim, S. H. (2016). Political polarization on twitter: Implications for the use of social media in digital governments. *Government Information Quarterly*, 33(4), 777-782. 10.1016/j.giq.2016.04.007.
- Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2020). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*, 119(1), 1-6. 10.1073/pnas.2025334119
- Lane, J. E., McCaffree, K., & Shults, F. L. (2021). Is radicalization reinforced by social media censorship? *ARXIV*, 1-16. 10.48550/arXiv.2103.12842
- Meeks, L. (2020). Defining the Enemy: How Donald Trump Frames the News Media. *Journalism & Mass Communication Quarterly*, 97(1), 211-234. 10.1177/1077699019857676
- Meta - FAQs. (2022). Meta Investor Relations FAQ. Retrieved June 17, 2022, from <https://investor.fb.com/resources/default.aspx#:~:text=Meta%20Investor%20Relations%3F,What%20is%20Meta's%20mission%20statement%3F,support%20and%20make%20a%20difference.>

- Comacho, M. (2019). *The era of post-truth, post-veracity and charlatanism*. Grupo Ciencia, Razón y Fe (CRYF). Universidad de Navarra. Retrieved June 17, 2022, from <https://www.unav.edu/web/ciencia-razon-y-fe/the-era-of-post-truth-post-veracity-and-charlatanism>
- Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2(2), 175-220. 1089-2680/98
- Political Polarization in the American Public*. (2014, June 12). Pew Research. Retrieved June 17, 2022, from <https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/>
- Ralston, S., Kliestik, T., Rowland, Z., & Vrbka, J. (2018). Are Pervasive Systems of Fake News Provision Sowing Confusion? The Role of Digital Media Platforms in the Production and Consumption of Factually Dubious Content. *Geopolitics, History, and International Relations*, 10(2), 30-36. 10.22381/GHIR10220183