

“Where Did Their Tweets Go?": A Quantitative Analysis of Parliamentarians' “Missing Tweets” in Western Europe

Joseph Noonan

Uppsala University

Department of Government

Autumn Semester 2021

Supervisor: Per Adman

Abstract

This thesis examines the factors behind politicians' 'missing tweets'. Missing tweets are posts that were once available on Twitter but are no longer accessible. Despite numerous studies on the Twitter behavior of politicians, few explore the dynamics around politicians' missing tweets. This study fills this research gap by examining the extent of, and possible factors associated with, missing tweets among parliamentarians active on Twitter in 2018 in six Western European countries (Denmark, France, Germany, Italy, the United Kingdom, and Sweden). Empirically, this study uses replication material from Castanho Silva and Proksch (2021a) to build a new dataset that identifies the number of missing tweets per parliamentarian, finding that overall, 21.8% of the tweets from 2018 had disappeared from the platform by October 2021. To determine the factors associated with these missing tweets four hypotheses are tested, examining mass deletion, gendered incivility, intra-party conflict, and populism. The results find that there is no association between gender or intra-party conflict and missing tweets. Furthermore, parliamentarians generally do not engage in mass deletion of tweets, but when they do those who have since left parliament are overrepresented. Lastly, there is a positive association between the level of populism and the number of missing tweets. The results of this thesis highlight both the theoretical and empirical importance of examining missing tweets when analyzing the behavior of politicians on Twitter.

Keywords: social media, Twitter, intra-party conflict, populism, gendered incivility, legislative politics

Revision history

This text has been revised from the version that was submitted and examined for the Master Thesis course at Uppsala University's Department of Government.¹ In this version, the appendices have been expanded with a lengthier discussion of the data collection procedure and the presentation of additional models with alternative specifications of the dependent variable.

¹ The submitted text is available on Digitala Vetenskapliga Arkivet (DiVA) platform: <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1628988&dswid=3254>

Acknowledgments

I wish to thank my supervisor Per Adman for his guidance in this process, encouraging me to investigate this topic, despite the unknowns and uncertainties surrounding it. I would also like to thank all of the staff at the Department of Government at Uppsala University, who have consistently nurtured my curiosity and interest in the study of political science. Thank you, Emelie, for your support and constant encouragement while putting up with me talking about my thesis all the time for four months. Finally, thank you to Jason, Teddy, and William for your years of friendship and support, and a special thanks to William for his critical comments on an earlier draft of this thesis.

Contents

Introduction.....	6
Outline.....	8
Background.....	8
The rights of politicians on Twitter	9
Theoretical background.....	11
The new transparency paradigm: Optimists and skeptics	12
Problematizing the right for politicians to be forgotten online	13
Conclusion.....	17
Previous research	17
Social media and Twitter	18
Social media as a data source	19
Previous research on missing posts on Twitter.....	20
Descriptive analysis of missing tweets	20
Political actors and missing tweets	20
Why the research gap?	22
Politicians' use of Twitter.....	23
Political parties and social media.....	23
Principal-agent theory of accountability.....	24
Populist parties and use of Twitter	27
Online incivility	29
Conclusion	30
Hypotheses.....	31
Mass deletion.....	31
Online incivility	32
Party dynamics	32
Intra-party conflict.....	32
Populist parties	32
Data and materials.....	33
Case selection	33
Generalizability	34
Data collection.....	34
Validation	35
Descriptive statistics	36
Tweet inequalities.....	37
Methods.....	38
Mass deletion.....	38

Regression analysis.....	39
Method and variable selection.....	39
Intra-party conflict.....	40
Populism.....	41
Results.....	42
H1: Mass deletion.....	42
H2: Gender.....	43
H3. Intra-party conflict.....	44
H4: Populist parties.....	46
Discussion.....	48
Mass deletion.....	48
Gendered incivility.....	48
Party effects.....	49
Intra-party conflict.....	49
Populist parties.....	50
Broader implications.....	51
Limitations and suggestions for future research.....	52
Limitations.....	52
Suggestions for future research.....	53
Conclusions.....	55
Bibliography.....	57
Appendix.....	63
A.1. Tweet Inequality.....	63
A.2. Mass Deletion.....	68
A.3. Regression analysis.....	69

Introduction

Social media has transformed communication. The ability to instantaneously communicate with millions of people, unbounded by geography with a low barrier to entry has created new opportunities and challenges for how humans communicate. For politicians, social media offers a way to communicate directly with the public, bypassing gatekeepers such as party leaders and the media. This new technology gives more freedom to politicians to decide who, what and how to communicate. At the same time, social media is visible and transparent, allowing citizens, party leaders, and the media to watch and engage with discussions in real-time providing a transparency tool that can facilitate accountability (Ceron, 2017a).

But social media does not just give politicians individualized control of who, what, and how to communicate it also gives them control to remove this communication. Despite the appearance of offering a complete archive of posts, social media is not static. With a click of a button, politicians can remove posts that they no longer want associated with them. Because of the networked nature of social media, posts that politicians have reshared (retweeted) that are later removed from the platform they will also disappear from the politicians' timeline. These actions create missing posts throughout the network. These are posts we know existed at one point on the platform but are no longer accessible. Any type of post removal creates ripples of these information gaps across social media. For example, when on January 8th, 2021, Twitter suspended President Donald J. Trump's personal Twitter account (@realDonaldTrump), for violating the glorification of violence policy, the historical record of all his tweets, and in turn, all his tweets that had been reshared by other users disappeared from the platform. With this suspension, the 26,000 tweets sent by Donald Trump and during his presidency and all the retweets by other users vanished, disappearing without an official archive (Forgey, 2021; Paul, 2021).

The suspension of President Donald J. Trump's Twitter account was a historic and highly publicized event. But posts going missing on social media is a regular occurrence, often going unnoticed, and are difficult to decipher through the online information flood. On Twitter, posts vanish at a roughly estimated rate of 10% per year (Barberá and Steinert-Threlkeld, 2020, p. 5). Recovering or archiving these Tweets is limited by data protection laws and in turn Twitter's terms of services (ToS) which prioritize users 'right to be forgotten' explicitly forbidding the

preservation of deleted Tweets that were earlier downloaded from the platform. In representative democracies, expanding the right to be forgotten online to politicians creates tension between the right to privacy and democratic norms of transparency. As transparency is a necessary component of accountability, these gaps in the social media record may be limiting the usefulness of social media as an accountability tool for voters, activists, journalists, civil society as well as political parties themselves.

Despite this new form of information control for politicians and the democratic implications of missing social media posts, there has been little academic research on this subject. This is surprising, as previous research has highlighted how social media allows for unmediated, individualized communication by politicians, which is sometimes at odds with party leaders and can indicate intra-party conflict (Castanho Silva and Proksch, 2021a; Ceron, 2017a, 2017b; Sältzer, 2020). Understanding why tweets are removed can shed new light on how political communication works on social media, both at the level of individual MPs and political parties.

This study aims to fill this research gap by conducting a quantitative analysis of the missing tweets of parliamentarians in six Western European countries (Denmark, France, Germany, Italy, the United Kingdom, and Sweden) on Twitter. Specifically, this study uses replication material from Castanho Silva and Proksch (2021a) to identify the number of tweets per parliamentarian that disappeared between 2018 and 2021. The term missing tweets is used throughout due to the limitations of gathering Twitter data programmatically, which does not distinguish if a user has deleted their post (active deletion) or if the post that is missing is a retweet of another user's post (network-effects). Missing tweets are tweets that we know existed at one point on Twitter but are no longer accessible on the platform. This research is novel with few identified studies examining missing social media posts among politicians and none in a comparative setting. Because there is limited identified research of politicians' missing tweets, this study is exploratory and multiple hypotheses are tested.

The overarching research question is as follows:

To what extent do parliamentarians active on Twitter in Western Europe have missing tweets, and what factors can explain these missing tweets?

By answering this question, this research contributes to the field in three ways. Firstly, deepens the discussion on the tension between transparency and privacy for political actors on social media. Secondly, it expands the literature on the behavior of politicians on social media by examining an overlooked research area. Lastly, it contributes to the broader discussion on the validity of social media as a data source by examining the potential biases of missing social media data in a specific subgroup.

Outline

This thesis begins with the section *Background* which presents an overview of how data protection laws and social media platforms' terms of service (ToS) together limit the ability of civil society, activists, and researchers to examine politicians deleted social media material. This section illustrates how this limits research on this topic and provides the necessary information to understand its democratic implications. *Theoretical background* explores the wider relevance of this field of inquiry to political science examining specifically the tradeoffs between digital privacy rights and the transparency of politicians on social media. This is followed by *Previous research*. This section outlines previous empirical and theoretical research relevant to the research question. It begins by examining previous research on Twitter in general, followed by previous studies on post-removal on Twitter, and lastly the behavior of politicians on social media. It is in this section that the justification for the hypotheses is derived. These four hypotheses are presented in the following section *Hypotheses*. The section *Data and materials* outline the case selection and the data gathering and validation procedures to create the novel dataset of the number of MPs missing tweets between 2018 and 2021, as well as descriptive statistics of this dataset. The methods used to test the hypothesis are then outlined in the *Methods* section results are presented in the *Results* section directly after and the broader implications of these results are discussed in the *Discussion* section. In the final section, *Limitations and suggestions for future research*, an overview of the limitations of this study as well as suggestions for future research is presented.

Background

This section provides a brief background of how both data protection laws and social media platforms' terms of service (ToS) constrain the ability to archive politicians' social media posts. This section provides background knowledge of both technical restrictions of examining missing

posts, which is essential to understanding the empirical section of this study as well as necessary background information to understand the wider democratic implications of these regulations.

The rights of politicians on Twitter

When a member of parliament opens a Twitter account, they do not need to declare that they are a politician. Outside of accounts associated with official communication channels of select states and candidates in American national elections, Twitter does not assign special labels to politicians (Twitter, 2021a, 2021b). Politicians are treated as ‘normal’ users and are therefore subject to the same policies and rights as non-public officials. This includes the right to data privacy. These rights are outlined in data handling legislation which in turn inform the contents of Twitter’s ToS.

For parliamentarians in Western Europe, much of data protection rights are informed by the General Data Protection Regulation (GDPR) which grants EU citizens and residence several rights with regard to how personal data is handled by organizations. These rights include among others the right to access, the right to rectification, the right to restriction of processing, and the right to erasure (or right to be forgotten). This allows for residents of the EU to request that data controllers (any organization that handles personal data) to remove their data from their platforms. However, GDPR recognizes that these rights are not absolute – stating that “the protection of personal data is not an absolute right” and that it should be “balanced in relation to other fundamental rights in society” (European Commission, 2016). The exceptions to the right to be forgotten highlight that the right of freedom of expression and freedom of information trump the right to be forgotten as well as when the information serves “the public interest” (Ibid). For example, GDPR clearly outlines that there is a journalistic exemption that trumps principles of privacy, for instance, a politician who says something controversial or inflammatory could not, for instance, invoke the right to be forgotten to remove their personal information and quotes from newspapers (Reventlow, 2020, p. 32).

As a result of these regulations, Twitter’s ToS requires that those using Application Programming Interface (API) services to programmatically download material must delete any data as soon as they are deleted on Twitter. Failure to do so can result in Twitter disabling one’s access to the API, although compliance with the ToS is not regularly checked by Twitter (Twitter, 2020; Uršič,

2016, pp. 4-5,7).² In practice, this means civil society groups, activists, voters, or researchers who programmatically download data from Twitter to their computers or databases are required to delete tweets in these archives as they are deleted on the platform (Twitter, 2020).³

These conditions severely limit both research on the nature of deleted tweets by politicians as well as transparency initiatives by watchdog groups such as Politwoops which archives politicians' deleted tweets. The case of Politwoops illustrates that social media platforms themselves struggle with the trade-off between privacy rights on one hand and transparency on the other. It also illustrates the precarious position for those examining missing tweets. The Politwoops project was originally developed by the Dutch Open States Foundation to increase the transparency and accountability of politicians by archiving and making politicians' removed tweets publicly available. This transparency initiative was adopted by civil society organizations in other countries, including the Sunlight Foundation in the United States. Because the archiving of deleted tweets violates the Twitter ToS, Twitter revoked Politwoops API access in the summer of 2015, citing the right to privacy was a priority for Twitter "whether the user is anonymous or a member of Congress" (Hern, 2015; Uršič, 2016, p. 5). The company's justification went further, describing how tweeting would be "terrifying" if tweets were allowed to be archived and that deleting tweets is a matter of freedom of expression (Hern, 2015). However, Twitter's stance of the universal right to privacy for all users of their platform was short-lived with Twitter reinstating Politwoops' API access several months later and publicly declaring their (new) commitment to "holding public officials accountable" (Crowell, 2015). This reversal of policy occurred after Twitter and Politwoops reached an agreement on API access, the details of which are not publicly known (Meeks, 2018, p. 7). As of 2021, Politwoops remains active, both in the United States and several other countries. However, somewhat ironically, the methodology of the

² APIs allow for researchers and other users to interact with social media platforms using a programming language. This allows users to for instance download information about specific users (such as the age of an account or how many followers they have). It also allows for users to download tweets in bulk, which contain both the text and metadata about the status (how many likes and retweets it has at the time the data was downloaded). As the data produced by social media is immense access to the API is essential to large scale data collection.

³ The full text of this stipulation read: "Removals. If Twitter Content is deleted, gains protected status, or is otherwise suspended, withheld, modified, or removed from the Twitter Applications (including removal of location information), you will make all reasonable efforts to delete or modify such Twitter Content (as applicable) as soon as possible, and in any case within 24 hours after a written request to do so by Twitter or by a Twitter user with regard to their Twitter Content, unless prohibited by applicable law or regulation and with the express written permission of Twitter." (Twitter, 2020)

databases is somewhat opaque, and it is unclear what the criteria for tracking candidates, politicians, and organizations are. As Meeks (2018) noted, there are significant gaps in who the database is tracking, so it is unclear how complete these databases are. Notably, several Politwoops affiliated Twitter accounts that repost the content of deleted tweets have been suspended, showcasing the possible continued reluctance of Twitter to allow this sort of archiving, or at least the precarity of civil society organizations that work in the grey area of tech platforms' ToS.⁴

Despite Twitter's fundamental rebalancing of privacy rights and transparency for politicians by moving from first defending the rights of privacy as absolute to switching to emphasize the importance of accountability the justifications why these rights were rebalanced remain opaque and unjustified. In the following section *Theoretical background*, the tension between balancing these rights is examined.

Theoretical background

As politicians' missing social media posts is a novel research subject, this section outlines the broader theoretical relevance of the research question in the field of political science. Specifically, it examines the tension between digital privacy rights on one hand and transparency and political accountability on the other, which the examination of missing material highlights. This has normative implications about how these rights should be balanced in liberal democracies. This section critically examines the role of transparency in modern democracies while in turn examining arguments for and against the right for politicians to be forgotten on social media, shedding light on the question: "*What are the democratic implications of missing social media posts?*"⁵

⁴ Of the country specific Twitter pages, the accounts for Iran, France, Italy, Switzerland, Catalonia, the European Union, and Argentina have been suspended. This is 30% of the total *Politwoops* country specific Twitter accounts. As Twitter does not provide specific justifications why accounts were removed so it is impossible to know with certainty the motivation behind these suspensions.

⁵ This discussion is concerned with the balancing of these rights in stable consolidated democracies. In weaker democracies, hybrid regimes and authoritarian regimes, any discussion of transparency and regulating digital privacy should consider the possibility that this regulatory infrastructure can be deployed for repressive purposes (see Tucker, J.A. et al. (2017), for a discussion on the asymmetry of hate speech laws in democratic and non-democratic settings).

The new transparency paradigm: Optimists and skeptics

The arguments for transparency in democratic systems stem from a logic that views transparency as a fundamental part of accountability, as observation is a necessary component to gain knowledge which then, in turn, can be used to sanction or reward public officials (Ananny and Crawford, 2018, p. 974; Baume and Papadopoulos, 2018, p. 175). This conception of political transparency was popularized by Enlightenment political theorists, particularly Jeremy Bentham who was a radical advocate for transparency arguing that “the more strictly we are watched, the better we behave” (Ananny and Crawford, 2018, p. 975; Gorwa and Ash, 2020, p. 288; Meijer, 2014, p. 507). Bentham further argued that transparency was not only an essential component for accountability, but also for increasing trust, legitimacy, and acceptance among citizens toward political leaders (Baume and Papadopoulos, 2018, p. 180). What transparency is and how it is realized, is a contested concept. Meijer (2014) offers a broad definition of transparency, defining it as “the availability of information about an actor allowing other actors to monitor the workings or performance of this actor”. Meijer (2014) further delineates the ways that transparency can be realized: passively (requiring freedom of information requests), pro-actively (publishing data), and forced access (through whistleblowing). Here, the archiving of deleted tweets would fall under the pro-active realization of transparency – in the case of politicians’ deleted tweets, this is currently being implemented by civil society groups with the tacit support of the online platform. While transparency is concerned with the availability of information, privacy can be conceptualized in terms of control of the availability of information, allowing individuals to maintain how information about oneself is used and communicated to others (Thompson, 2011, p. 60). These two concepts are inherently in tension with one another, as they are both concerned with the availability of information, but each emphasizes the importance of control in opposite areas, with transparency emphasizing control of information for those monitoring actors and privacy emphasizing an individual’s control over information about themselves.

In liberal conceptions of democracy, which view political institutions both as a safeguard and a threat to individual rights, checks, and balances to hold public officials accountable are especially important (Skaaning, 2021, p. 31). Here, the appeal of transparency for public officials would seem to be self-evident, as for checks and balances to work, voters and other actors who hold public officials accountable must know how they are performing. However, the embrace of transparency as a democratic good is a relatively recent phenomenon, with most democracies

only introducing freedom of information legislation in the 1970s and the 1980s (Meijer, 2014, p. 509). For example, in the United States, whose governmental structure embraces a liberal conception of democracy with robust checks and balances, it wasn't until 1970 that congressional representatives' votes on amendments to bills were public records (Schudson, 2020, p. 1671). This is a remarkably different situation today, where the expansion of transparency initiatives over the last fifty years has led transparency to become an essential feature of modern democracies (Meijer, 2014, p. 507). Today, transparency has entered the standard vocabulary of democratic norms, alongside other concepts like freedom of speech and free and fair elections.

With the rise of this transparency paradigm, there is a predictable backlash of scholarly criticism of the transparency ideal and its sometimes lofty promises with scholars calling the embrace of transparency “a new religion”, “a magic concept” and “a panacea”(Baume and Papadopoulos, 2018, p. 187; Meijer, 2014, p. 507,521). It is often ill-defined and is not problematized, as it represents an essential democratic ‘good’. Ananny and Crawford (2018) argue that transparency also can be characterized as not just a way of clarifying or making information available but also an observation system that promises control. In the case of archiving social media posts by politicians, it is hoped that this will lead to some form of good communicative behavior. However, it is unclear if this in practice has occurred. This raises one of the major criticism of the promises of transparency, that transparency without institutional arrangements that foster accountability, can be disconnected from power and thus has no meaningful effect (Ananny and Crawford, 2018, p. 187). Furthermore, the early assumptions of transparency having a positive effect on trust may have been overoptimistic, as information with empirical evidence shows that transparency only sometimes increases trust (Baume and Papadopoulos, 2018, pp. 180–181). Other scholars argue that not only does the promise of fostering trust fail, but the basis of radical transparency comes from a cynical view towards power creating a “culture of suspicion” that itself can erode trust (Meijer, 2014, p. 519). Lastly, others argue that transparency can have negative effects among public officials as it decreases honest communication and can harm deliberation (Schudson, 2020, pp. 1675–1676).

[Problematizing the right for politicians to be forgotten online](#)

Despite the increasing embrace of transparency in democratic societies, the regulations concerning deleted posts on social media prioritize privacy for public officials by treating posts on Twitter as personal data rather than public statements (Uršič, 2016, p. 5). Here it is important

to note, that, in contrast to many discussions about transparency, archiving social media posts does not always provide insight into a decision-making system. Social media is instead communicative material that is already crafted to a public audience (Golbeck et al., 2010, pp. 1619–1620). Here, three arguments can be made about the importance of this material despite being “just” communicative.

Firstly, and most importantly, political communication in representative democracy has an integral role in the accountability of politicians to voters as politicians communicate their policy preferences and promises of action to citizens (Ceron, 2017a, p. 205). Without this communication, citizens would be unable to set a benchmark of what to expect of their representative and thus their ability to punish or reward politicians and political parties that did not live up to the electoral promise. As elections are spaced out years apart from each other, the historical record of past social media posts may be even more relevant for voters to use as a benchmarking tool than more recent posts. This also implies a temporality, where preserving politicians’ social media material is justified during the electoral cycle, which in some cases, means preservation before they are in a decision-making position.

Secondly, through all public communicative processes, in both traditional and social media, politicians engage in impression management, which consists of all activities in which an individual monitors information that is presented about them to control the impressions that this information makes on others (Leary and Kowalski, 1990, p. 34; Meeks, 2018, pp. 2–3). Thus, the concepts of privacy and impression management are linked as they are both concerned with an individual’s control of information about oneself. With social media, politicians have gained a new powerful tool of impression management which is unmediated by traditional gatekeepers, who have a ‘watch dog’ role where they had control of statements made by politicians in different contexts and can use this information to inform the public of the actions of public officials. Digital privacy rights can enhance impression management by politicians as in the case of removed tweets, the right to not have tweets archived allows politicians to control information by hiding content consistent with undesirable images (Meeks, 2018, p. 3). Because social media is already less restrained than traditional media, with politicians in control of their content, a tweet that is deleted because of its substantive content would indicate an opinion or view that they would prefer to be hidden. Thus, citizens and other actors who hold politicians accountable

get better insight into the ‘true’ character, motivations, and preferences of individual politicians (Castanho Silva and Proksch, 2021a, p. 2). Lastly, social media posts can provide insight into intra-party conflict which often takes place behind closed doors (Ceron, 2017a, p. 20). By preserving this material voters gain a deeper insight into the ideological factions and functioning of political parties.

For radical transparency advocates like Bentham, the distinction between communicative material and material concerned with decision making and implementation of policy would not be of high importance, as he would argue that all manner of secrecy increases anxiety among the public and leads to decreased trust and legitimacy among the citizenry (Baume and Papadopoulos, 2018, p. 180). This perspective on transparency is imbued with a sense of cynicism, that any attempt at ‘hiding’ information is nefarious. The example of Politwoops echoes this ideology with their justification for their service writing that the platform allows users to “*Explore the tweets they would prefer you couldn’t see*”(Politwoops, n.d.).

For scholars with a more skeptical view of transparency, the tracking and archiving politicians deleted tweets, no matter how innocuous they may be, may erode trust in public officials by emphasizing the perceived attempts at secrecy without respect for the content thus eroding trust without substantive reasons. Digitalized transparency allows for massive amounts of information to be shared. Here, the volume of both social media posts and those that have been deleted create an impossible amount of information to digest. This deluge of information provided by digital transparency initiatives can create an information overload that hurts rather than improves citizens' decision-making (Baume and Papadopoulos, 2018, p. 181; Etzioni, 2010, p. 402; Meijer, 2014, p. 512). At its most extreme, this combination of a cynical view of power combined with an excess of information can confuse rather than inform the public (Meijer, 2014, p. 520). For instance, during the 2016 American presidential election, the organization Wikileaks released 20,000 pages of emails of then-candidate Hillary Clinton’s campaign chair John Podesta (Bleakley, 2021, pp. 8–9). The release of these emails, often devoid of context, was fused with misinformation on social media which lead to the emergence of wide-reaching anti-system conspiracy theories about the political elite that birthed the QAnon movement (Ibid). While this is an extreme example, it exemplifies the problems that many transparency skeptics raise, that the assumption that observation does not increase truth and that transparency can decrease trust,

while too much information can harm rather than help decision making (Baume and Papadopoulos, 2018, p. 181).

One of the complications with preserving deleted social media data is that this does not merely represent the professional or public positions of public officials, but a personalized account that documents a rather wide spectrum of activities that blur the lines between private and public life which allows voters to ‘peak behind the scenes’ (Jacobs and Spierings, 2016, p. 16). This personalization can be further broken down into professional personalization, which emphasizes the politician over the party unit, emotional personalization, and lastly sharing private information about themselves (Metz et al., 2020, p. 1483). For privacy concerns, the sharing of private information is perhaps most relevant. In this case, politicians are not only presenting their ideological positions but also presenting themselves in a personalized manner, showcasing parts of their personal lives such as their family or other aspects of their home life. This, however, cannot be construed solely as a peak into the personal lives of politicians but as an active communication strategy to enhance “relatability”, which strategically emphasizes parts of their private lives as a strategy of impression management (Meeks, 2018, p. 2; Thompson, 2011, p. 58). The entanglement of the personal deployed as a professional communication strategy creates difficulties in a broad justification of limiting politicians' right to disappear on social media. If for instance, exceptions to the right to be forgotten for politicians were implanted, this would allow actors to preserve elements of the politicians’ personal lives, which implies that those personal aspects of politicians’ lives are “in the public interest” despite having questionable democratic utility.

These privacy concerns are further exacerbated because of the reciprocal nature of social media which means that politicians can engage with normal users. For instance, politicians can ‘retweet’ and force any user on the platform into the spotlight (Meeks, 2018, p. 9). If the user then later deletes the original post that was retweeted and this data is allowed to be preserved, this becomes a violation of the rights of privacy to citizens who politicians engage with on social media (*Ibid.*). The aspect of reciprocity is a fundamental part of social media, which makes it distinct from traditional media (Jacobs and Spierings, 2016, pp. 62–63). Furthermore, it is often pointed to as one of the democratically positive effects of social media allowing politicians to move beyond top-down communication style to engage directly with constituents creating a

virtuous feedback loop of trust (Tromble, 2018a, p. 692, 2018b, p. 236). Although perhaps hyperbolic to call the archiving of tweets a ‘terrifying’ idea as Twitter did, there is a possibility that there is a chilling effect of excessive archiving. This may lead politicians to revert to a top-down communication style which in turn would decrease reciprocity – although the effect of this might be minimal due to the already public nature of tweets.

Lastly, arguing that politicians have a lesser claim to the right to be forgotten online leads to temporal questions as to when this applies and when it still serves “the public interest”. For instance, should it only be those who hold political office, or should this also include after an individual leaves a political position? The archiving of material after a politician leaves office may become increasingly relevant in the European political context where there is an increasing amount of interest in the democratic implications of post-parliamentary careers and the ‘revolving door’ of public servants moving into the private or lobbying sector (Rasmussen et al., 2021, p. 488).

Conclusion

This section outlined and problematized transparency norms by bringing attention to different perspectives brought forth by scholars. Specifically, regarding the archiving of deleted social media material by politicians, there seems to be an unexamined consensus within the (limited) literature that there should be exceptions for these actors, that it is both democratically and ethically justifiable to examine this material (Freelon, 2021, p. 8; Meeks, 2018, p. 11). This is in line with a broader shift towards a norm of transparency for public officials. However, as this section outlined, there are many implicit tradeoffs with this perspective that should be weighed against each other when making normative claims about archiving deleted material.

Previous research

This section outlines both previous empirical and theoretical studies relevant to the research question. The section is organized as follows: First, a brief outline of what social media is and the platform characteristics of Twitter are presented. This discussion is essential in understanding how post-removal works on these platforms, which is central to the focus of this study’s research. After this, the use of social media data in social science research is outlined. This is followed by a review of the limited research on post-removal behavior on Twitter, both by ‘normal’ users and political actors. Lastly, research on politicians and the use of Twitter is presented. This section

concludes with a presentation of four questions that guide the hypotheses formation. These questions are derived from both the limited literature on Twitter post removal and the behavior of politicians on social media.

Social media and Twitter

Social media generally refers to online platforms where users themselves can generate, organize, and access content from a self-selected network of other users all of which are updating constantly (Klinger and Svensson, 2015, p. 1245). Different platforms have different ‘digital architectures’ which enable or constrain user behavior online (Bossetta, 2018). Twitter’s architecture is characterized by its microblogging nature where users create short posts (tweets) which are character limited (first to 140 characters, which was then doubled to 280 in 2017). This character limitation increases the number of posts made compared to other social media platforms. Users can follow other users (referred to as friends) with a stream of all tweets from friends appearing in a constantly updating timeline, which appears in reverse chronological order (Bossetta, 2018, p. 487). This architecture emphasizes the present with the founder Jack Dorsey originally envisioning a platform that discouraged retroactive reflection (Weller, 2013, p. xxii). Users can interact with each other in several ways that appear publicly. Users can directly address other users by posting a tweet and tagging their handle with the at-sign (for example @BarackObama), like tweets, repost another user’s tweet (retweeting) or quote other users’ tweet to comment on their tweet (a feature added in 2015). Connectivity on the platform is unidirectional by default (one does not need to consent to be followed) and user accounts are open by default so one does not need to be following someone to see their tweets (Bossetta, 2018, p. 479). Twitter users tend to be younger, more male, have higher levels of education, and are more politically attentive, while those engaging in political discussion were mostly men, living in urban areas with stronger ideological positions (Barberá and Rivero, 2015; Mellon and Prosser, 2017; Mislove et al., 2021; Wojcik and Hughes, 2019). Among politicians, there is a widespread adaptation of Twitter (Barberá and Zeitzoff, 2018; Castanho Silva and Proksch, 2021a, p. 5; Jungherr, 2016). Because of this userbase, Twitter is often considered an ‘elite platform’ (Jacobs et al., 2020, p. 614).

These characteristics of Twitter make it an ideal platform to analyze missing social media posts by politicians. Firstly, Twitter encourages a high volume of posting due to structural limitations

on the number of characters. Secondly, communication on Twitter is highly public as by default it does not require user consent to be followed. Lastly, the platform has attracted an ‘elite’ audience with high levels of interest in politics and the platform has been widely adopted by politicians.

Social media as a data source

Social media offers an unprecedented amount of easily accessible data for social science research. Barberá and Steinert-Threlkeld (2020) provide a concise outline of the benefits and pitfalls of this form of data. Using APIs researchers can easily download rich datasets of real-world human behavior in an online environment. They note that, in contrast to other observational datasets, this data isn’t generated in a lab or survey setting, removing the risk for social desirability bias. For cost-constrained researchers, this offers an obvious advantage compared to costly field studies and other forms of data generation (Ibid. p. 3).

For this study, an important aspect of Twitter data is that Twitter is a living dataset. The data that is collected is a snapshot of when the data was collected not when the data was generated. This is an important distinction because, while the textual data remains unchanged if it is still accessible, the quantitative data associated with both the account (for instance the number of followers and friends) and the tweet itself (the number of likes and retweets) can be constantly shifting (Zubiaga, 2018, p. 982). This means that data downloaded directly from when a tweet is posted can change if the same data is downloaded later. As users are not allowed to share full Twitter datasets, restricted to only sharing unique identifiers that are later used to repopulate the dataset, this creates issues with the reproducibility of Twitter research (Ibid., p. 975).

Furthermore, and key to this study, if a user’s account is removed, or their account is set to private or deletes their tweets, then this data is no longer accessible anywhere on the platform (Almuhimedi et al., 2013, p. 989; Maddock et al., 2015, p. 3). If a post is removed that has been retweeted, the tweet is also removed from the user who reshared the post. The effect of posts no longer being accessible is known as post-rot or absent-data’ (Barberá and Steinert-Threlkeld, 2020, p. 5; Freelon, 2021, p. 7). In this study, the term ‘missing tweets’ is used for clarity. While this missingness is a known issue with Twitter data the literature on the characteristics is limited and even more so when it comes to political topics.

Previous research on missing posts on Twitter

Descriptive analysis of missing tweets

Most of the research conducted on missing tweets on Twitter consists of descriptive studies from the field of computational science without grounding in social theory. Almuhimedi, et al., (2013) is one of the first studies to examine missing posts on Twitter of randomly selected users and found that almost two thirds (65.2%) of tweets were deleted within an hour and of the deleted tweets 17% were deleted for superficial reasons such as correcting phrasing or fixing typos. Liu et al (2014) expands on these findings with a larger dataset, collected over a longer period, and found that 20% of tweets were removed in their sample, with 10% of the removals stemming from users' privacy settings and 5% from explicit deletions. Zubiaga (2018) provides the most comprehensive overview examining 30 different Twitter datasets of general users collected during specific events from 2012 to 2016. This study found that in total 18.6% of tweets were missing, steadily increasing as the datasets became older, with some datasets missing up to 30% of their Tweets (Ibid. p. 982). Lastly, Joan Timoneda (2018) contributed to this literature by examining the patterns of tweet removal across thirty-six different topics instead of users, finding that the removal rates by topics yield interesting and somewhat contradictory results, with tweets about politicians having a lower removal rate than trivial topics, while tweets about political events having higher removal rates.

Crucially these studies focused on the persistence of general Twitter datasets in general, including trolls and spam accounts that are subject to active removal by Twitter. Despite these differences from the focus of this study, this descriptive literature shows that post-removal is widespread on Twitter and that there are indications that removal patterns are non-random.

Political actors and missing tweets

Despite the issue of missing tweets on Twitter, very few studies have specifically researched missing tweets by political actors. Meeks (2018) provides the only research article exploring politicians missing tweets. The author focuses on examining the possibility of using the *Politwoops* database to examine politicians' deleted tweets. Meeks (2018) argues the importance of studying these tweets, noting how the deleted tweets can inform research on impression management research, which is broadly defined as research with focuses on self-presentation or all the activities which serve to influence other participants. For politicians, this is hyper-relevant as they live highly public lives with competitive elections where impressions of the candidate are

very important. The advantage of using deleted tweets in this area of study is that it allows for analysis of not just their desired image but also how undesirable images are avoided.

Methodologically, Meeks (2018) advocates for an intertextual qualitative approach for this type of analysis, where data is collected over a given period, and each of the deleted tweets being analyzed within its wider context, for instance by taking into consideration whether deleted tweets are simply reposted with typos corrected or rephrasing. The intertextual approach is important for understanding the impact or strategic choices that go into rephrasing that would capture nuances better than a quantitative decontextualized approach. Unfortunately, due to the lack of textual data for the missing tweets, this study cannot use this method and there have been no identified articles that have conducted research using this framework or research exploring why political actors delete tweets.

Ringel and Davidson (2020) conduct one of the few studies that attempt to move beyond a non-theoretical descriptive analysis of missing tweets by examining why people delete tweets. As such, this thesis draws heavily from this study in hypotheses formation. In their study, they examine the motivations behind Twitter deletion behavior among journalists. The authors conducted 14 in-depth interviews with New York City-based journalists that actively use Twitter. They found that the journalists that they interviewed deliberately and actively engaged in tweet removal. Many only conducted surgical removals of tweets, for instance searching their timeline for questionable words that might reflect poorly on them. However, many of those interviewed also used third-party apps to mass delete tweets. This was done by either removing all previous tweets, “nuking them”, or by using the same software to remove tweets during a certain time span, for instance keeping tweets up only for one month. This allows the tweets to ‘expire’ after a set amount of time. These mass deletion acts leave tremendous gaps in their Twitter archive. The authors dub behavior where users intentionally remove their content as “proactive ephemerality”.

Ringel and Davidson (2020) found two key motivations for journalists to delete tweets: harassment and occupational risks. The journalist interviewed regularly faced harassment on Twitter a platform that was described as easily “weaponized” fostering a “gotcha culture” (Ibid. p. 8). Having experienced harassment on Twitter, interviewees saw deleting tweets as a preemptive tool to avoid future harassment. The authors found a stark gendered component in how harassment was experienced, with female journalists fearing for their safety to a higher

degree than male journalists, and as such embrace deletion practices more readily. In terms of occupational hazards, journalists face an unstable labor market and were worried that editors or hiring managers would scrutinize their Twitter feeds resulting in reputational risk. One notable example is of an unnamed journalist who was widely criticized for deleting all their tweets before accepting a political advisory role. While some of the interviewees saw that this behavior was a violation of transparency norms in journalism, most respondents saw this as legitimate practice as they viewed Twitter as an essentially ephemeral platform, where tweets are made in a particular moment and are more similar to a verbal conversation. In general, the risk of safety and professional stability trumped more ideological motivations as a commitment to transparency.

Although focusing on public agencies, McCammon's (2020) study can provide some further insights into the reasons why political actors remove tweets. Using freedom of information requests McCammon, asked for records and justifications for tweet deletions. Many agencies provided little or no justification for the deletion behavior of tweets, stating a lack of records and referring to Twitter's lack of archival tweets. Of the few agencies who justified deleted tweets included, many indicated that tweet deletions were due to minor content issues such as bad images or typos. However, in one example of a controversial tweet by the Pentagon, the deletion itself was seen as a potential vulnerability as the press drew more attention to the tweet after it was deleted.

While both McCammon's (2020) and Ringel and Davidson (2020) shed some light on how actors within a political system (journalists and public agencies), and can perhaps be generalized to some extent, there is still a research gap on the behavior of politicians, with no articles examining the patterns or the reasons to why politicians do (or do not) delete tweets.

Why the research gap?

The limitation in the study of deleted tweets is that the Twitter ToS for use of their API explicitly forbids the storing of deleted tweets if the user knows that the tweet has been removed, switched to protected status, or is suspended (Twitter, 2020). Thus, without explicit permission from Twitter, researchers are limited to an existing dataset that has reached agreements with Twitter (such as Politwoops), breaking the Twitter ToS, or following the ToS by only removing textual and metadata from deleted tweets resulting in binary data to flag whether a tweet has been

missing or not. This lack of textual data and metadata of deleted tweets creates methodological difficulties in the analysis of deleted tweets.

Politicians' use of Twitter

Jacobs and Spiergers (2016) offer a theoretical framework that differentiates social media from traditional media and how that impacts politicians' use of social media. Compared to traditional media, social media is unmediated, personal, and interactive, with low barriers of entry, and high possibilities of virality. These characteristics in turn create four opportunities for politicians. By allowing users to directly post individual messages thus creating an *advertisement opportunity* to reach a wide audience. These messages can be further tailored to certain communities allowing for a *target-group opportunity*. Furthermore, Twitter allows politicians to create a personal connection by allowing the audience to 'peek behind the curtain' of their day-to-day lives and have reciprocal communication at a low barrier to entry. This creates *human contact opportunities*. Furthermore, especially on elite platforms like Twitter, social media opens a *salon-debate opportunity* by allowing politicians to contact journalists and other elite actors and debate them directly. Lastly, there is a potential for virality in a network-based media sphere, which further enhances the *advertisement opportunity* of social media.

This framework can be complemented by Castanho Silva and Proksch (2021a), who make a distinction between *amplifying* and *substituting* for differentiating how politicians can use Twitter for different purposes to exploit these opportunities. Twitter can act as an *amplifier* of messaging for the party that is already present in other areas. Twitter can also act as a *substitute*, where the unmediated nature of social media allows politicians to bypass gatekeepers, speak in an unconstrained way, and communicate beyond the party brand.

Political parties and social media

The dynamics around these opportunities are impacted by the party system of the country being studied. Much of the earlier work on politicians and social media use was dominated by research in the United States, which is unique in its extremely personalized, first-past-the-post electoral system with liberal campaign finance laws (Jacobs and Spierings, 2016, p. 5). In this system, candidates' individual competition for seats creates a clear appeal for the personalized communication style of social media, but in party-dominated systems, which are prevalent in Western Europe, the role personalized role of social media may be contradicted by traditional party-centric communication strategies, in which the party apparatus asserts more control over

communication both online and offline (Enli and Skogerbø, 2013, pp. 757–758). These parties work actively to maintain party discipline and avoid politicians using social media as a substitute for the message of the party.

Principal-agent theory of accountability

A principal-agent theory of accountability can provide a useful framework to understand the dynamics of sanctioning or rewarding politicians for online communication and how this shapes parliamentarians' behavior. Unlike the discussion in the *Theoretical background*, this discussion of accountability is focused on shedding light on the *behavior* of parliamentarians rather than the broader democratic implications of transparency and accountability. In short, the principal-agent theory is a type of rational choice modeling that conceptualizes accountability as a process where agents are obligated to act on behalf of principals while these principals can sanction, or reward actors based on their performance due to formal or informal institutional arrangements (Fearon, 1999, p. 55; Gailmard, 2014, p. 50). This is a highly generalizable model that can be applied across different organizations. Here, it is important to delineate between what Mark Philp (2009) terms *formal* and *political* accountability. *Formal* accountability is concerned with the regulation of the legitimate use of power whereas *political* accountability is concerned with whether principals would approve or endorse the actions taken. Unless otherwise mentioned, accountability in this study is referring to *political* accountability.

In the *Theoretical background* section, the discussion of accountability in democracies emphasizes the role of voters as the main principals of politicians with regular elections acting as a formal institution that allows sanctioning or rewarding. This is a simplified model of accountability, which ignores different principals that constrain or sanction politicians during and in-between elections. In the case of legislators, they have multiple competing principals in addition to voters, such as party leaders, presidents, governors, political activists, interest groups, and donors (Carey, 2008, p. 4). For this study, the principal-agent conceptualization of accountability is particularly useful, because it acknowledges that parliamentarians have multiple principals who move them in often contrary directions. Because principals require information about what the agents are doing to sanction or reward agents, transparency is a necessary criterion for accountability, although transparency alone is not sufficient in assuring accountability (Ibid). These sanctioning or rewarding behaviors can occur between electoral periods, excreting pressure during the day-to-day operations of politics (Ceron, 2017a, p. 11).

These pressures, in turn, determine the level of party (dis)unity, with increases in the number of competing principals increasing the level of intra-party conflict. With this model accountability and intra-party conflict are linked as there is an inherent tension between individual accountability which requires MPs to be directly responsive to a diverse set of interests and collective accountability which requires parties to engage in cohesive collective actions to pursue policy goals and in turn increase re-election prospects (Castanho Silva and Proksch, 2021a, p. 1; Ceron, 2017b, p. 8; Sältzer, 2020, p. 2).

The rise of social media has led to questions about how this new media form affects the accountability of legislators to their principals. Ceron (2017a) theorized that social media platforms may create a new principal of online public opinion, which would be responsive to the concerns and pressures of this mixed group of stakeholders. However, his research showed that legislators were unresponsive to social media pressure, suggesting that these platforms service not as new competing principals, but rather as a low-cost tool for existing principals to monitor legislators (Ceron, 2017a, p. 208). Here, the tensions between *individual accountability* and *collective accountability* come to a head, with voters and activists able to pressure individual legislators on issues using social media while party leaders can use the same tools to observe parliamentarians and to enforce party discipline, discouraging *substituting* communicative behavior that goes against the party line while encouraging *amplifying* communicative behavior. In party-dominate systems, intra-party conflict can often be hidden behind party doors so the ability to monitor the preferences of individual politicians can lead to a higher level of transparency of the political system, especially for principals such as voters that are outside of the party (Ceron, 2017a, pp. 19–20).

Jacobs and Spierings (2016) provide examples of how parties try to monitor and control social media, in other words, how parties try to enforce accountability on their MPs by encouraging *amplifying* behavior. The authors found that parties take three approaches to try and control the social media output of MPs: forbidding or discouraging social media use, monitoring and controlling, and training and supporting social media use (Jacobs and Spierings, 2016, pp. 107–111). They found that almost all parties utilized the monitor and control approach, in which MPs' social media presence was constantly monitored by social media managers who would reprehend them if they tweeted too far out of the party line (Jacobs and Spierings, 2016, p. 111). The

prevalence of this approach further strengthens Ceron's (2017a) conceptualization of social media as a low-cost transparency tool for different principals.

However, despite these attempts at control, social media remains less mediated by party or media gatekeepers than traditional communication channels. Furthermore, deviation from the party line in communication may result in fewer consequences from the parties in general in comparison to deviation in voting (Sältzer, 2020, p. 3). Thus, the views expressed by politicians may be closer to the actual preferences of an individual politician. Because of this social media has been used as a measurement of the policy preferences of individual politicians (Barberá, 2015). This has been particularly fruitful for measuring intra-party conflicts. Previously, scholars of intra-party conflict and party unity such as Carey (2008) have used legislative votes to measure (dis)unity and consider the visibility of the votes by legislators to constitute an essential component of individual accountability. However, voting or formal documents as a source of intra-party conflict may be hiding intra-party tensions as some conflicts are not always formally represented in official documents nor through parliamentary activity and parties are incentivized to present cohesive preferences to voters (Ceron, 2017b, p. 8; Sältzer, 2020, p. 2). In Ceron (2017a) the authors utilized automated text analysis to analyze the online posts of Italian politicians and found that it successfully predicted politicians leaving parties, being appointed ministers, and endorsement of candidates. While Sältzer (2020) employed a similar approach to the MPs in Germany and found that while their tweets reflected party positions and dimensions, members of factions were closer to other members of their factions on cultural and economic dimensions. Lastly, Castanho Silva and Proksch (2021a) combined Twitter data with records of legislative speeches, finding that while sentiment about the EU on Twitter reflected the party position there was evidence of intra-party dissent with MPs using Twitter as a *substitution* channel. Those who used Twitter as a *substitution* channel, whose individual positions were farther away from their party spoke less in parliamentary speeches.

Overlooked in the literature on the intra-party dynamics of Twitter communication is the ability for MPs to remove tweets. This allows an even higher level of individualization of communication, as individual MPs can not only decide what and who to communicate but they also gain the power to remove posts. This, in turn, may limit the power of principals to use social

media as a monitoring channel, as MPs have the power to selectively remove posts to avoid sanctioning by party elites.

Populist parties and use of Twitter

In addition to intra-party conflict, the ideological base of a given party may also affect how social media is used. Here, the dynamic of populist parties and social media present an interesting case of how ideology and party structure impact online behavior as these two elements are often at odds with each other. Populist party ideology emphasizes closeness to ‘the people’ opposing them to traditional elites while party structure on the other hand which is often highly centralized and intolerant of dissent (Jacobs and Spierings, 2019, p. 1683). Definitions and conceptualizations of populism are diffuse and debated, however, the *ideational* approach to populism has become the most popular today. This approach conceptualizes populism as being a ‘thin’ ideology characterized by anti-elitism that separates and contrasts ‘the good people’ with ‘corrupted elites’ and argues that politics should be an expression of the general will of the people (Ernst et al., 2017, p. 1348; Jacobs and Spierings, 2019, p. 1683; Mudde, 2017, pp. 27–30; Stanyer et al., 2016). Because this approach views populism as a thin ideology that latches onto a thicker ideology populist parties can be left and right-wing, with left-wing populists using some form of socialism as an ideological basis and right-wing populists using neoliberalism or nationalism as theirs (Mudde, 2017, p. 38). Radical right-wing (RRW) populism can be distinguished primarily by the core focus on nativism, that is the belief that membership to the nation-state should be exclusive to a “native group” while expelling “non-native” elements, as well embrace of traditional values, accusations of elites sacrificing the interests of the nation for internationalism (Bar-On, 2018, pp. 53–54; Rydgren, 2007, p. 242).

For studies that are examining political communication, it is important to clarify the conceptual framework used to define populism. Populism is sometimes defined as a political communication *style* that can be employed by a wide range of political actors (Stanyer et al., 2016, p. 354). Here it is *communication* rather than political actors that are defined as populist (Ibid.). Thus, this conception allows for a more continuous scale depending on the frequency or infrequency of this communication style, which is characterized by simplistic, negatively charged emotional attacks on elites while rhetorically positioning the political actor as the true representative of ‘the people’ (de Vreese et al., 2018, p. 426; Ernst et al., 2017, p. 1350; Jacobs et al., 2020, p. 613). This differs from an actor-centered approach which uses *a priori* characteristics about the parties to

classify politicians as populist or non-populist and then analyze the differences in their communicative behavior (Jacobs et al., 2020, pp. 612–613; Stanyer et al., 2016, p. 354). In this text, unless otherwise noted, populism is conceptualized using the actor-centered approach. This is for two reasons, firstly the *a priori* conceptualization is most suitable for examining how populist parties differ from non-populist parties in how they use social media and communication in general (Jacobs et al., 2020, p. 615; Stanyer et al., 2016, p. 354). Secondly, a communicative approach for this study would require access to the textual data of the missing tweets which is not available in this study.

While the nature of social media would seem to fit ideologically with populist parties as they emphasize the closeness to ‘the people’ it does not appear that populist politicians embrace social media more than other parties. In Jacobs and Spierings’ (2019) study on the adaption and use of Twitter by Dutch MPs, they found that MPs belonging to populist parties were less likely to adopt Twitter, engaged in fewer reciprocal interactions and befriended fewer people, have a higher tendency to retweet members of their network but tweeted a similar amount as other parties. The authors suggest that this may be due to the structures of these parties, as they are often more hierarchical, thus despite having ideological reasons to embrace these media platforms, the party structure informs their behavior on Twitter. However, their findings are limited due to the number of populist parties in the Netherlands, which raises questions about how well this finding generalizes to other countries.

Of particular interest to this study is the role of RRW populists. The contents of RRW populist politicians' posts on social media have higher levels of anti-immigrant and anti-elitist references on social media and citizens who follow RRW populist politicians on social media have higher levels of anti-immigrant attitudes (Heiss and Matthes, 2020). Furthermore, these parties more frequently name and shame journalists creating hostile interactions (Jacobs et al., 2020, p. 624). However, RRW populist politicians tend to be less extreme in their language than their followers, using more neutral language and preferring implicit rather than explicit forms of ostracizing out-groups (Engesser et al., 2017, p. 1119). This may indicate both ban-avoidance behaviors on the platforms and to comply with hate speech laws (Åkerlund, 2020, p. 623; Engesser et al., 2017, p. 1119). Despite this seeming self-moderation to avoid legal action or platform removal, discursively some parties, such as Alternative für Deutschland (AfD), share many discursive

frames with more extreme radical right political movements (Ahmed and Psoiu, 2021). Right wing populist politicians also have a higher prevalence of bot activity within their social networks both in terms of followers and friends (Castanho Silva and Proksch, 2021b). Some right wing populist parties, such as AfD, have openly admitted to using bots as communication strategy on social media (Keller and Klinger, 2019, p. 176). In short, these parties' online Twitter presence is characterized by embeddedness in twitter networks with high levels of potentially rule-breaking tweets and users. These characteristics increase the possibility that these users retweet material that would later be removed actively by Twitter or by posting material themselves that violates Twitter rules on hate speech.

Online incivility

Ringel and Davidson (2020) found that a key motivation for the deletion of tweets was a pre-emptive strategy to avoid harassment. This motivation was particularly strong among female journalists. Like journalists, politicians also are regularly exposed to incivility on Twitter, although there is a wide variation in the estimated volume of incivility. Theocharis et al. (2020) estimate that 18% of all tweets mentioning American legislators to be uncivil, as defined as “disrespectful discourse that silences or denigrates alternative views” (Theocharis et al., 2020, p. 3). This is a broad definition that includes using words like “arrogant” to highly offensive slurs. It does not distinguish between political incivility and more extreme political intolerance (Southern and Harmer, 2021, p. 261). In the highly polarized context of the US, they find that there is a baseline of incivility that spikes during political contentious moments and that although organized harassment campaigns exist, the bulk of incivility is from a heterogeneous group of users. Moving to a European setting, Ward and McLoughlin (2020) examine the prevalence of abuse and hate speech directed towards MPs in the UK. They define abuse as unwanted contact whose intent is to cause harm, and hate speech as “hatred towards a group based on protected characteristics”. They found that 2.57 % of tweets were classified as abusive, with a wide variation between MPs. In total 62% of MPs have had at least one abusive tweet sent to them in the period studied. The prevalence of hate speech was lower with less than 0.42% of tweets containing hate speech. Like Theocharis et al. (2020), the authors also found that abusive tweets increased during specific political events.

Due to numerous high-profile gendered attacks against MPs online, the gendered effects of online incivility and intolerance have been widely studied (Southern and Harmer, 2021, pp. 259–260).

Theocharis et al. (2020) did not find a significant gendered effect of incivility, although the lack of this effect may be a measurement artifact. This is in line with Ward and McLoughlin (2020), who found that female MPs had lower levels of abuse compared to men. However, because Ward and McLoughlin (2020), have a finer delineation between types of incivil content, measuring both abuse and hate speech, they can measure the difference in the severity of the type of incivil content. They find that women MPs have higher levels of hate speech which denotes a higher level of threatening behavior. This complements Southern and Harmer's (2021) study which found that there were slight gendered differences in terms of volume of incivility but larger differences in the content of incivility with women MPs having higher levels of identity-based incivility. Furthermore, high-profile cases of online harassment, and the lived experience of harassment in other realms of political life, may lead female politicians to self-censor or participate less on online platforms (Southern and Harmer, 2021, p. 272; Ward and McLoughlin, 2020, p. 65).

Conclusion

This section has provided an overview of the relevant literature on social media post removals and parliamentarians' behavior on Twitter. From this review of the existing literature four specific questions, related to the overarching research question, have been formulated:

1. *Do MPs actively remove their tweets through mass deletion?*
2. *Does gendered incivility affect MPs tweet removal patterns?*
3. *Does intra-party conflict affect MPs tweet removal patterns?*
4. *What is the effect of populist ideology on missing tweets?*

Hypotheses

This section outlines both the hypotheses on missing tweets and the underlying justifications for each of the given hypotheses. Broadly speaking, these hypotheses are broken into two levels: the MP level and the party level. MP level effects are characteristics that are specific to the individual politician, whereas the party level effects focus on how party dynamics constrain or encourage behavior that affects the level of missing tweets.

There are two main groups of mechanisms that cause tweets to be removed. Firstly, there is active removal, where a user deletes their tweet, sets their account to private, or removes their account. Secondly, there are network effects, which are downstream effects from when a user retweets a post that is later actively removed, by the original user. For example, if user A posts a tweet that is then retweeted by user B and then later user A deletes their original tweet, the retweet by user B will also be removed. This is an important distinction, as one requires active participation whereas the other is an artifact of who the politician retweets.

Mass deletion

Ringel and Davidson (2020) found that journalists regularly engaged in active removal of their tweets to mitigate occupational risks including harassment and the potentiality of previous tweets reflecting poorly on them in the future. Like journalists, politicians may view their historical tweets as a liability due to scrutiny from various principals and may use third-party software to mass delete their tweets. As social media can be conceived as a tool of transparency for those in principal actor relationships, the strategic removal of tweets may be a method to minimize accountability to different principals. As Ringel and Davidson (2020) was a qualitative study examining journalists, it is difficult to estimate how their findings will translate to parliamentarians. However, given this strong norm in a specific context of Ringel and Davidson (2020), it is reasonable to expect that at least in some subgroups of MPs there is a norm of deletion of tweets using third-party automated tools.

This hypothesis is unique in this study because it examines the prevalence of a certain type of removal tool rather than correlative associations with the MPs background characteristic or party and the number of missing tweets. Thus, unlike the other hypothesis, the findings resulting from this hypothesis can infer a degree of intentionality in removal, as it is highly unlikely that total removal of historical tweets is the result of networked effects.

H1: Parliamentarians widely engage in mass deletion of their historical tweets

Online incivility

Ringel and Davidson (2020) noted that harassment was a strong motivation to delete tweets and this harassment was gendered with female journalists experiencing harsher forms of harassment. The gendered nature of online harassment has also been observed among politicians with Ward and McLouglin (2020) and Southern Harmer (2021), finding online incivility higher among female politicians. Because of the impact of this incivility, H2 theorizes that female MPs may have more missing tweets, as a mitigation measure to buffer against online incivility and hate speech. This hypothesis argues that the potential causal mechanism is the unobserved hostility in the replies to female MPs' tweets. As the textual data for the replies to deleted tweets are unavailable any associations would require further research to isolate the correlation between the level of hostility to replies of deleted tweets and active removal.

H2: Female parliamentarians have more missing tweets than male parliamentarians.

Party dynamics

Intra-party conflict

Due to active monitoring of social media feeds by party elites, MPs who are ideologically out of line with their party and use Twitter as a *substitution* tool to communicate free of partisan constraints may engage in higher levels of tweet removal to avoid accountability by party principals who seek to maintain an *amplifying* communication strategy close to the party brand to maintain collective cohesion. Here the logic of tweet removal echoes Ringel and Davidson's (2020) finding that journalists removed tweets due to fear of scrutiny by their principals, such as editors and future employers.

H3. MPs who use Twitter as a substitution tool will have more missing tweets than those who use it as an amplifying tool

Populist parties

Populist parties are characterized by a contradiction between an ideology that emphasizes a closeness with the people and party structure which often encourages top-down discipline. This hierarchical structure may encourage MPs to remove tweets proactively to avoid sanctioning by party elites (active removal). Twitter behavior for RRW populist politicians specifically is characterized by dense twitter networks, retweeting each other members, with higher levels of bot

activity and misinformation, and sharing content that has high levels of anti-immigrant messaging, which may violate Twitter rules. This leads to a higher risk of retweeting or posting material from accounts that will later be removed (network effects). Concerning populist parties, three hypotheses are derived:

H4a: MPs in parties with higher levels of populism will have more missing tweets than MPs of other parties.

H4b: Party structure is more explanatory than populist ideology in explaining MPs missing tweets.

H4c. The effect of populism on missing tweets is mediated by radical right-wing ideology.

Data and materials

This section explains the case selection procedure, the data gathering procedure, and the validation of the data. Due to the nature of Twitter as a ‘living dataset’, the difficulty of working with large quantities of data and the novelty of the data it is important to verify that the data gathered is robust and error-free. This discussion is followed by simple descriptive statistics of the dataset.

Case selection

As this research requires the availability of the large-scale longitudinal Twitter data of parliamentarians there is limited flexibility with case selection. The replication data from Castanho Silva and Proksch (2021a) provide one of the few data sources possible for this study. This replication data provides lists of tweet IDs and usernames of MPs in seven countries (Denmark, France, Germany, Italy, Spain, Sweden, and the UK) that were collected during 2018 as well as an aggregated dataset that includes relevant descriptive data on the MPs. The unit of observation are MPs with Twitter accounts. For this study, the country selection is relevant for two reasons. Firstly, MPs in these countries have adopted Twitter at high rates, ranging from 66% in Italy to 95% in France (Castanho Silva and Proksch, 2021a, p. 4). Secondly, these countries are all consolidated democracies. In weak democratic regimes or autocracies where there is a higher risk of severe punishment or harassment which may alter online behavior in general and behavior around post-removals. As all hypotheses about online behavior have been derived from research in democratic contexts, this case selection is justified. Furthermore, this sample of countries

provides a diversity of electoral systems and media systems, with two countries having list proportional representation system (Denmark and Sweden), two having a mixed-member proportional system (Germany and Italy), and two having plurality systems (France and the UK) (Lijphart, 2012, p. 133). This diversity provides increased generalizability to these hypotheses which are concerned with MP level and party level factors.

In contrast to Castanho Silva and Proksch (2021a), this study does not use data from Spain. As the Spanish sample is split into two legislative sessions due to a cabinet change in June 2018, the data provided by Castanho Silva and Proksch (2021a) splits the Spanish sample into two periods. As this splits the sample into an earlier period and a late period, this may create unknown bias as the networked effects of tweet removal have a temporal dimension with older tweets having a higher rate of missingness (see Zubiaga (2018)). If the temporal effects were known and could be controlled for this split sample could offer an interesting insight into how cabinet reshuffles affect removal patterns.

Generalizability

Due to the diverse selection of countries, and the nearly complete coverage of parliamentarians' tweets it is likely that the findings of this study generalize to other Western Democracies with high Twitter uptake among parliamentarians in 2018. However, it may not generalize in other time periods as the architecture of Twitter is constantly shifting. For instance, the dynamics of harassment and tweet deletion may have changed as Twitter has begun to offer more tools to prevent harassment, such as limiting who can respond to tweets (Xie, 2020). This may have eliminated or reduced the need to delete tweets to avoid the effects of harassment. There may be other more subtle changes to the architecture since then which shifts deletion behavior in unobserved ways.

Data collection

The dataset has been built using replication data from Castanho Silva and Proksch (2021a). In the original dataset, the authors captured these tweets using the *streamR* package for R which allows the capture of tweets in real-time. The authors captured the tweets in real-time from February 16th to December 31st, 2018, and then used a historical search to capture tweets from January 1st, 2018. Twitter's ToS only allows tweet IDs and screen names can be shared for replication so the full textual content and metadata attached to it cannot be shared. However, because the data is

linked to MPs' usernames, it is possible to still get counts of the number of tweets posted (total tweet IDs) by MPs in the original sample.

To build the dataset of missing tweets, the original Castanho Silva and Proksch (2021a) dataset has been 'rehydrated' by redownloading the tweets using the *rtweet* package for R, which allows easy access to the Twitter API (Kearney, 2019).⁶ The final set of tweets was downloaded between October 19th and October 20th, 2021. This represents the 2021 sample, which is all the tweets still tweets from 2018 still available in 2021. The Twitter API does not provide notification if a Tweet has been deleted, rather it simply does not return any data for a given Tweet ID. After the tweets were downloaded, the total number per username was counted for both the 2018 and 2021 samples. The number of missing tweets is simply the total tweets in 2021 minus the total tweets in 2018. Because the metadata of the original tweet is unavailable, it is impossible to disaggregate between original tweets and retweeted tweets. Thus, active removal or network effects cannot be independently tested.

In some cases, the status ID was reused by clearly unrelated accounts after the status in question had been removed, these tweets had different usernames and therefore were not included in the tally of existing tweets by MP. A further complication in the data collection process is that Twitter allows users to change their screen names. To accommodate this, all accounts in the 2021 sample, that were not matched in the 2018 samples but had over two tweets, were manually matched with their respective 2018 screen names.⁷

Validation

To check that these tweets were missing, a random sample of 1,000 of the missing tweets was manually checked to ensure that they were removed. Although Twitter does not provide information on a tweet's deletion status through API request, when a user tries to access a deleted

⁶ A detailed technical explanation of this process is available in the section of the annex *A.1. Detailed technical explanation of data gathering and validation procedure*

⁷ Screen names were validated by looking at the new screen names Twitter biography. Most of the changes were small often removing mentions of parliamentary status (removing MP suffix), adding, or changing party affiliation, or other cosmetic changes in the screen name (for exempling changing from nooshiit to dadgostarnooshi). Three accounts were removed from the dataset. Boris Johnson was removed from the dataset. This is because the account being tracked was not his personal account but the account of the Foreign Office, of which he was the secretary at the time. French MP Valérie Oppelt and German MP Michael Grosse-Brömer were also removed as their screen names changed during the period of study, thus the dataset included duplicates for these MPs.

tweet through a web browser a message will appear indicating that the tweet has been deleted, is from an account that no longer exists, is protected, or is from a suspended account. To expedite the process the *webshots* package for R was used to take screenshots of all of the sampled Tweets (Chang, 2019). In some cases, the Tweet IDs had been reused meaning that Tweets were redirected to other users' tweets. These were coded as removed and redirected if the tweet was clearly not made by the MP in question.⁸

In the sample, 96.9% of the flagged tweets were deleted, 2.7% of the flagged tweets were not available due to changes in user privacy, and 0.66% were posted by accounts that were since deleted. Of the tweets that were removed, 3.3% were removed due to the account being associated with them being suspended.⁹ This process verified that the tweets flagged as removed are missing from the platform, thus it is valid to classify them as missing tweets.

Descriptive statistics

Table 1. provides some basic descriptive statistics of the dataset. In the entire sample, the total percent of missing tweets is 21.8%. The percentage of missing tweets per country ranges from 28.2% in Germany to 19.1% in the UK. There are no clear patterns between missing tweets and the country's electoral systems. The rate of missing tweets seems to roughly correspond with previous research with Barberá and Steinert-Threlkeld (2020) reporting anecdotal evidence of post disappearance at a rate of 10% per year and Zubiaga (2018) who found 18.6% of tweets were missing in their 30-dataset sample. Although these are rough comparisons due to differences in temporal span and users sampled. As this study's sample only includes verified parliamentarians, it is surprising that the missingness rate is similar to samples of the general Twitter userbase which is more susceptible to spam, trolls, and other rule-breaking behavior that might increase the amount of missing tweets.

⁸ Generally, it was evident that these tweets were not made by the MP, for instance promotional material for companies in languages not commonly spoken in the EU. If there was any ambiguity, coding was determined by closer inspection of the user's profile

⁹ As there is no way to identify if these tweets are original tweets that were removed by the MP in question, or simply retweets by users who then deleted their tweet the finding that 3.3% of tweets that were removed because their users had been suspended should not be extrapolated to assume that politicians were retweeting accounts that were at the time violating Twitter ToS. It could be that these accounts began violating terms of service after politicians retweeted them either through a change in behavior or due to the user being hacked.

Table 1. Descriptive statistics by country

Country	Total Tweets in 2018	Tweets Remaining in 2021	% Missing	Num. Parties	MPs on Twitter
Denmark	65,383	50,314	23%	9	140
France	381,209	292,974	23.1%	8	510
Germany	233,016	166,210	28%	7	455
Italy	176,779	140,627	20.5%	5	372
Sweden	132,736	101,627	23.4%	8	272
UK	730,530	590,646	19.1%	7	502

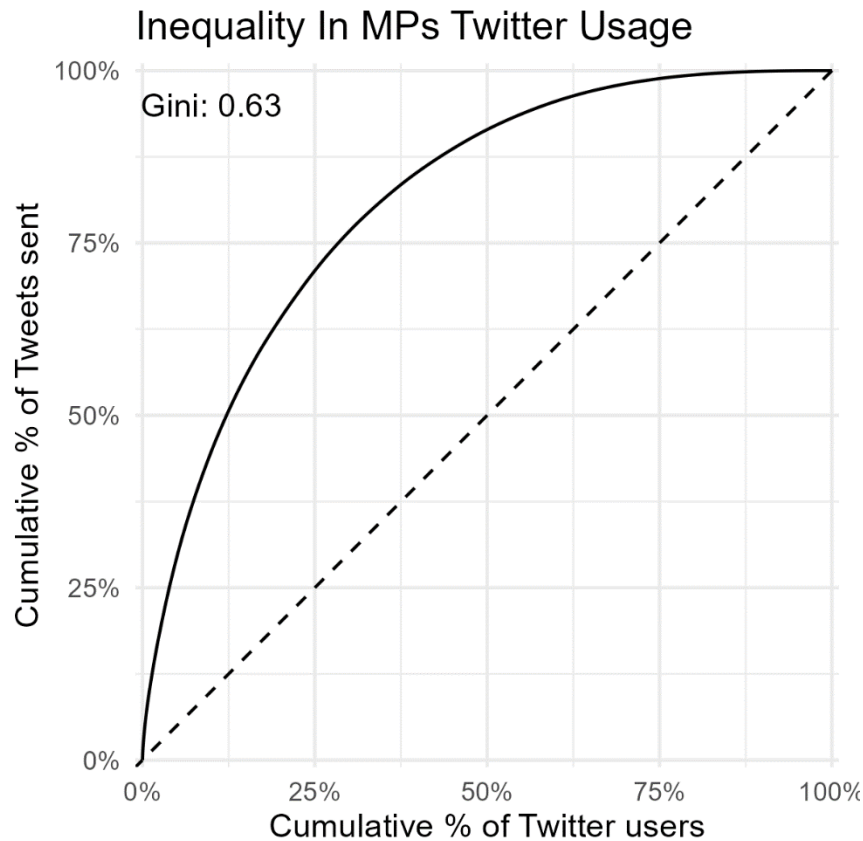
Tweet inequalities

The production of tweets is highly unequal with a small subset of users producing the majority of the tweets while many other users tweet very little (Barberá and Rivero, 2015, pp. 714–715).

Here measurements normally used for income inequality, the Gini coefficient, and Lorenz graphs can be useful in understanding the distribution of tweets and removed tweets (Barberá and Rivero, 2015; Theocharis et al., 2020). The Gini coefficient runs from a scale of 0 to 1 with zero being perfect equality in distribution and 1 being total inequality. In this dataset, the overall Gini coefficient of tweets in 2018 is 0.63 while the coefficient for missing tweets in 2021 is 0.64.

However, this distribution varies from country to country, ranging from the lowest levels of inequality (0.49) in France to the highest (0.71) in Sweden. Figure 1. Shows the Lorenz curve of MPs Tweet activity in 2018. In less abstract terms, this distribution of tweets means that the top 1% of MP tweeters post 24% of the tweets, the top 10% produces 53%, and the top 25% post 72% of all tweets.

Figure 1. Inequalities in MPs Twitter Usage



Tweets in this sample appear to be more equal than tweets in the general public, with previous research finding a Gini coefficient in the US at between 0.75 and 0.77, while in Spain which is more relevant for this study, the Gini has been estimated at 0.76 (Barberá and Rivero, 2015, p. 715; Theocharis et al., 2020, p. 11).

Methods

Mass deletion

As outlined in Ringel and Davidson (2020), there are many external web applications available that allow for automated post deletion. These programs can be used to mass delete all of a users' tweets or delete tweets after a certain period allowing them to 'expire'. Utilizing these tools would leave substantial gaps in the historical record of MPs tweets. While there is ambiguity in whether individual missing tweets are done by the user or the results of network-based effects, mass deletion patterns leave substantial holes in the archive and would indicate that the missingness was the result of intentional deletions by the user.

To test this if members of parliament engage in mass deletion first a simple binary variable has been created coding MPs into two groups, those who have no remaining tweets accessible from 2018 on Twitter and those who do. Those who had no remaining tweets accessible and had at or over 100 tweets in 2018 were then manually coded to determine if the account was deleted, suspended, protected, or remaining. The cutoff of 100 was chosen to filter out users who were in the bottom 25% of tweet frequency indicating low levels of tweet activity, which wouldn't be consistent with the practice of mass removal tweets, which consisted of active users grooming their social media presence (Ringel and Davidson, 2020). Coding was conducted by visiting the Twitter page for each account and coding the account status message provided by Twitter. Three status messages have been coded. Accounts with the status message "This account doesn't exist" are coded as removed accounts. Accounts that have the status message "Account suspended" are coded as removed and suspended, while accounts that are protected were coded as protected. Users who engage in mass deletion would both have to have their account remaining after having removed all their tweets during the time studied.

Regression analysis

Method and variable selection

Hypothesis 2 through 4 are tested using Ordinary Least Square (OLS) regressions. This method is chosen, as it allows controls for the total number of tweets in 2018. As the number of missing tweets is highly dependent on the number of sent this control is essential for meaningful analysis. For each of these models, the dependent variable is the natural log of the number of missing tweets.¹⁰ This has been log-transformed due to the highly skewed nature of tweets (see Figure 1.). For each of the hypotheses tested, several control variables are introduced in the regression models. Every model presented includes the logged number of tweets in 2018 as a control. Subsequent models introduce controls at the MP level and the party level. The MP level control variables added are the number of terms the MP has served, whether the MP has ever held a ministerial position, and whether the MP is currently holding a leadership position. These control variables have been included for all hypotheses as the seniority of a member of parliament affects

¹⁰ In regression models with a log dependent variable the correlation coefficient of non-log independent variables is interpreted as the percentage increase in the dependent variable, while the log independent variable coefficients are interpreted as elasticity (Békés and Kézdi, 2021, p. 207; Nyman, 2021, p. 26). As logs cannot take zero values, a $\log(x+1)$ was used. An alternate specification of the dependent variable using the share of tweets remaining is presented in the appendix.

the amount of media attention that they receive, as well as differences in how they use social media platforms, with leaders and more senior officials adopting social media earlier, but using social media more as an amplifying tool, with less reciprocity with other members (Castanho Silva and Proksch, 2021a, p. 13; Jacobs and Spierings, 2019, p. 1689; Tromble, 2018a, p. 696). Party level controls have been added for the size of the party, whether the party is in government, and ideology are added. Party size and whether the party is in government have been included as they both are proxies for both being part of the political establishment and the number of resources the party has at its disposal (Tromble, 2018a, p. 696). All models include country fixed effects have been included in all models to account for cross-national differences in the Twitter environment and the political environment at large. Additionally, robust standard errors are calculated at the party level, to avoid underestimating the standard errors due to similarity within of MPs within parties.

The control variables at the MP level come from Castanho Silva and Proksch (2021a) while the size of the party and whether the party is in government comes from the ParlGov dataset, which Castanho Silva and Proksch (2021a) have merged into their replication data. The ideology variable comes from the Populist and Political Parties Expert Survey (POPPA) (Meijers and Zaslove, 2020, 2021). The ideology variable has been scaled at the party level to help with interpretation with zero representing the mean values and one unit increase corresponding to one standard deviation. Models are run using the *lm_robust()* function from the *estimatr* package (Blair et al., 2021).

Intra-party conflict

To test the effect of MPs using Twitter as a *substitution* channel, sentiment data for EU-related tweets from Castanho Silva and Proksch (2021a) is utilized. These aggregated variables can shed some light on the textual content of the tweets even among those that are now missing, as these data provide sentiment analysis scores for each MP, based on the Tweets that existed when the data was originally collected. Castanho Silva and Proksch (2021a) chose to operationalize *substitution* using EU sentiment, because it is salient in all countries in their sample, with clear stances for or against deeper integration and are a source of much intra-party division (Ibid. p. 2).

The variable *distance to average party EU sentiment* provides a measurement of the MPs sentiment difference in EU related tweets from the average of their respective party this is

complemented with *Distance to party (non-EU)* which measures non-EU related semantic difference from the rest of the respective parties. Castanho Silva and Proksch (2021a) calculate sentiment by using automatic translations of the text of tweets and then using a sentiment dictionary to assign positive and negative values to each word then the logged ratio of positive to negative terms is calculated. For both variables measuring the distance to party is the absolute difference between the MPs average sentiment score and their parties. This is scaled as a measurement of disunity, with 0 indicating total sentimental alignment between the MP and their party. Thus, the differential effects of negative or positive differences from the party are not tested with these variables.

Populism

To operationalize populism, this analysis uses the POPPA dataset (Meijers and Zaslove, 2020, 2021). The POPPA expert survey was fielded in Spring 2018, coinciding with the collection of the original Twitter data. This expert survey dataset takes a disaggregated approach to measure populism, arguing that it should be measured as a latent concept that is constitutive of different items. Using the ideational approach to populism the concept is broken down into five items: a Manichean worldview, the homogeneity of ‘ordinary people, the general will of the people, people-centrism, and anti-elitism. The dataset also includes measurements of populism as a political style or an organizational style, including questions on common sense and emotional appeals as well as the level of personalized leadership and intra-party democracy. All the questions are asked on an 11-point scale (0–10) resulting in continuous measurement of all the items. The advantage of a continuous scale is that it allows for more variation in the constitutive dimensions of the concept being measured which is advantageous when approaching the classification of borderline cases (Meijers and Zaslove, 2021, pp. 278–279). The *populism* variable provided by the POPPA dataset is the factor regression scores of the variables *manichean*, *indivisible*, *generalwill*, *peoplecentrism*, and *antielitism*. To test the effects of hierarchical party structure, the variable for intra-party democracy is used (*intradem*).¹¹ Like the ideology variable, the variables *populism* and *intradem*

¹¹ Question wording: Some political parties practice more intra-party democracy than others (i.e. party members play a role in decision making, there is room for internal debate, decision-making is inclusive of various factions and organizational layers within the party). Please tick the box that best describes each party's practice of intra-party democracy.

have been centered around the mean to ease interpretation. Lastly, to test the effects of radical right-wing populism, the variable *nativism* has been used as an interaction term with populism. Nativism was chosen as an interaction term as it is recognized as a fundamental component of RRW populism that does generally present itself in left-wing populist parties (Bar-On, 2018, p. 54; Meijers and Zaslove, 2021, p. 397).

Results

H1: Mass deletion

Table 2. shows the results of this analysis. In total, only 4.17% of MPs Twitter accounts had all their 2018 tweets missing, of those with more than 100 tweets in the 2018 sample, 77.2% of these accounts were no longer on Twitter, while 3 accounts (6.7%) were suspended and 1 (2.2%) had switched their privacy settings to ‘protected’ thus making it impossible to access their tweets. 18% (n = 8) of the accounts had removed all their content from 2018 but were still using the Twitter service, indicating mass deletion. This corresponds to 0.35% of the total sample of 2,251 MPs. These findings indicate that there is no norm of mass deletion among the MPs in the studies sample. Thus, the hypothesis that MPs engage in mass deletion can be rejected. This suggests, that unlike the journalists profiled in Ringel and Davidson (2020) there is not a norm of mass deletion among MPs in Western Europe, neither as a whole nor in specific subgroups such as specific countries or political parties.

During this coding exercise, it became apparent that many of the MPs with deleted accounts had left politics. Because of this, a new coding variable was added to flag if MPs had left or announced that they were going to leave parliament before the 2021 tweet sample was collected. Over half (52.2%) of MPs with no remaining tweets accessible had left politics between 2018 and 2021. While the scope of this study does not allow time for coding the total sample in this way, it is safe to assume that this rate of leaving parliament is an overrepresentation compared to the total sample.

Table 2. H1: Mass deletion results

Among full sample	
No remaining tweets are accessible	3.95% (n = 89)
Tweets still accessible	96% (n = 2,162)
Among no remaining tweets accessible accounts, with over 100 tweets in 2018 (n = 44)	
Deleted accounts	77.2% (n = 34)
Account remaining but tweets removed <i>Engaged in mass deletion</i>	20.5% (n = 9)
Suspended account	6.8% (n = 3)
Protected account	2.2% (n = 1)
Left parliament	52.2% (n = 23)

H2: Gender

The association between gender and missing tweets is tested through three regression models. Three models have been conducted with each model increasing the number of control variables. Model 1 includes only gender and the logged number of tweets in 2018, which is used to control for the frequency of Twitter use. Model 2 introduces further demographic control variables at the MP level while Model 3 introduces party controls. In addition to testing the main hypothesis, these models can also provide insight into other demographic and party factors that may affect the number of missing tweets.

Table 3. Gender results

DV: # of missing tweets (log)						
	M1: Gender		M2: MP controls		M3: MP and party controls	
Variable	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²
Male	0.04	0.034	0.04	0.034	0.03	0.034
N tweets in 2018 (log)	0.91****	0.014	0.91****	0.014	0.92****	0.015
Minister (ever)			-0.06	0.043	-0.08*	0.044
Leadership position			-0.06	0.085	-0.05	0.086
Terms in parliament			0.02*	0.012	0.03*	0.013
Party in government					0.08	0.061
Share of seats held by party					0.01	0.184
Party ideology left to right (scaled)					0.00	0.030
N	2,098		2,098		2,098	
Adjusted R ²	0.869		0.869		0.869	

¹*p<0.1; **p<0.05; ***p<0.01; ****p<0.001

²SE = Standard Error

Table 3. Shows the results of the regressions. Fixed effects at the country level are included but not reported. The results of every model show that gender has no significant effect on the number of tweets missing. The direction of the insignificant effect is pointed in the opposite of the hypothesized direction. The only control variable that reaches significance at the p<0.05 level is the logged number of tweets in 2018. At the p<0.1 level. In both Model 2 and Model 3 there is an association with the number of terms served and the number of missing tweets that is significant at the p<0.1 level. In Model 2 each additional term is associated with a 2% increase while in Model 3 each term is associated with a 3% increase. In Model 3, having ever served in a ministerial position is associated with an 8% reduction in the number of missing tweets which is significant at the p<0.1 level.

H3. Intra-party conflict

To test the effects of MPs using Twitter as a *substitution* channel on missing tweets three regressions have been run. Table 4. Presents these results. First, the regression results from a with minimal controls are presented with only *EU Dissent* on Twitter and a control for the logged

number of tweets in 2018. The next model presented has MP and party level control variables, while the final model includes *Distance to party overall (non-EU)* included to control for the general semantic difference between the MP and the rest of the party.

Table 4. Intra-party conflict results

DV: # of missing tweets (log)						
Variable	M1: Difference to Party (EU)		M2: Difference to Party (EU) Controls		M3: Diff to Party (Tone Controlled)	
	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²
EU Dissent on Twitter	0.07*	0.037	0.05	0.036	0.04	0.035
Distance to party on Twitter (non-EU)					0.17*	0.085
N tweets in 2018 (log)	1.0****	0.021	1.0****	0.021	1.0****	0.023
Terms in parliament			0.03**	0.012	0.02**	0.012
Male			0.03	0.034	0.03	0.034
Minister (ever)			-0.06	0.047	-0.06	0.047
Leadership position			-0.07	0.082	-0.06	0.081
Party in government			0.04	0.062	0.03	0.060
Share of seats held by party			0.02	0.166	0.04	0.168
Party ideology left to right (scaled)			0.03	0.029	0.03	0.028
N	1,862		1,862		1,862	
Adjusted R ²	0.841		0.842		0.843	

¹*p<0.1; **p<0.05; ***p<0.01; ****p<0.001

²SE = Standard Error

Model 1 shows a possible positive relationship between MPs semantic distance from their party on the EU tweets and missing tweets that significant at the p<0.1 level. With this association, each unit increase in semantic distance from the party on EU tweets is associated with a 7% increase in the number of missing tweets. However, the significance of this association disappears in Model 2 when MP and party controls are added. In Model 3, the semantic distance

between the MP and the party for all non-EU tweets has been added. This control is significant at the $p < 0.1$ level, with each unit increase in general semantic distance associated with a 17% increase in the level of missing tweets.

H4: Populist parties

To test the relation of populist party ideology on the missing tweets five different regression models have been run. Table 4. shows the results of these regressions. Model 1 through 4 focus on populism in general, while Model 5 introduces an interaction term between nativism and populism. In Models 1 through 4, populism is associated with an increase in the number of missing tweets. In Models 1, 3, and 4 this is significant at the $p < 0.05$ level, while in Model 2, which controls for only intra-party democracy and the number of tweets in 2018, this is only significant at the $p < 0.1$ level. The effect of an increase of one standard deviation of populism is associated with a 5% increase in the number of missing tweets in Model 1 and Model 2, and a 6% increase in Model 3 and Model 4. In Model 5 an interaction term between *nativism* and *populism* is added to determine if the relationship between populism and missing tweets is conditional on the party embracing nativism (the core element of RRW populist parties). However, there is no significant effect with this interaction term.

Table 5. Populism results

DV: # of missing tweets (log)										
Variable	M1: Populism		M2: Populism and intra-party democracy		M3: Populism with controls		M4: Populism and intra-party democracy with controls		M5: Populism x nativism Interaction	
	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²
Populism (scaled)	0.05**	0.023	0.05*	0.026	0.06**	0.025	0.06**	0.026	0.01	0.071
Intra-party democracy (scaled)			-0.01	0.023			0.01	0.032		
Populism X Nativism									0.08	0.092
Nativism (scaled)									-0.03	0.060
N tweets in 2018 (log)	0.91****	0.015	0.91****	0.015	0.92****	0.016	0.92****	0.016	0.92****	0.016
Terms in parliament					0.03**	0.013	0.03**	0.013	0.03**	0.013
Male					0.02	0.034	0.02	0.033	0.02	0.033
Minister (ever)					-0.07	0.045	-0.07	0.045	-0.07	0.044
Leadership position					-0.07	0.090	-0.07	0.090	-0.07	0.091
Party in government					0.09	0.056	0.09	0.056	0.10	0.063
Share of seats held by party					0.05	0.165	0.07	0.167	0.08	0.179
Party ideology left to right (scaled)					-0.01	0.025	-0.01	0.027	-0.02	0.041
N	2,094		2,094		2,094		2,094		2,094	
Adjusted R ²	0.869		0.869		0.870		0.870		0.870	

¹*p<0.1; **p<0.05; ***p<0.01; ****p<0.001

²SE = Standard Error

Discussion

Mass deletion

There is no evidence that parliamentarians commonly practice mass deletion of tweets. Only 0.35% of the total sample of 2,254 MPs showed evidence of proactively removing all their tweets while continuing to use Twitter. This contrasts with the norm of mass deletion of tweets practiced among journalists found in Ringel and Davidson (2020). In forming the hypothesis there were high levels of uncertainty as the only study looking at motivations of deleting tweets among political actors was a qualitative study of journalists. Thus, this result may be due to differing communication incentives and logics between journalists and members of parliament. A potential reason for this disparity is that there could be higher levels of scrutiny on members of parliament compared to individual journalists. As such it may be more difficult for politicians to mass remove tweets discretely. As mentioned, both in Ringel and Davidson (2020) and McCammon (2020) the deletion of tweets themselves can create negative press, in some cases perhaps generating more attention than the original post. This may be a modern reflection of Jeremy Bentham's assertion that "suspicion always attaches to mystery" (Baume and Papadopoulos, 2018, p. 180). Additionally, politicians, like the public at large, may already view tweets as ephemeral, something that appears in timelines but is rarely examined closely after the fact (Mikal et al., 2016, p. 4). Although this seems less likely, both due to the norm among journalists as found by Ringel and Davidson (2020) as well as the importance and awareness of impression management and communication for politicians (Meeks, 2018, pp. 2–3). However, this finding does not preclude the possibility that politicians engage in more surgical removal of past tweets, removing tweets that may expose vulnerabilities in the future while avoiding the negative attention that mass removal might entail. Lastly, the analysis showed that over half (52.2%) of the MPs who had no remaining tweets accessible had left politics since 2018. This suggests that deleting one's Twitter account may be done by ex-parliamentarians to exit the public sphere, and perhaps, like the journalists in Ringel and Davidson (2020), as a way to avoid scrutiny by future employers.

Gendered incivility

There is no evidence to support H2, that female MPs had more missing tweets compared to male MPs. The models do not show a significant effect of gender in either direction. This suggests that

if the gendered effects of online incivility result in more deletion, the effect is not substantial enough to be seen using the current measurement. This may be due to the relatively “low” volume of estimated hate speech and incivil speech relative to civil speech. It may be that the deletion of tweets as a harassment defense mechanism only occurs when there is “dogpiling” occurring when there is an organized or unorganized stream of directed harassment. Furthermore, as noted by Ward and McLoughlin (2020) and Southern and Harmer (2021), there may be a gendered difference in how women and male MPs post on Twitter, with the explicit intent to reduce the possibility of harassment.

Despite this null result, the control variables used indicated several interesting findings. Notably, having ever served in a ministerial position is associated with an 8% reduction in the missing tweets significant at the $p < 0.1$ level in the model with both MP and party controls. This may be an indication of the communication style that party elites adopt, favoring top-down amplifying communication styles over reciprocal conversations, which may lead to a decreased frequency of retweeting others outside of the party apparatus leading to fewer missing tweets due to network effects. These results should be interpreted with caution both due to the lower level of significance as well as the lack of significance of this variable in other models in this study. Additionally, in both models, each additional term is associated with an increase in the number of missing tweets which is significant at the $p < 0.1$ level. This finding is in line with the high number of account deletions by parliamentarians who had left office found in H1. Taken together these may suggest that as parliamentarians leave politics, they remove more posts on their social media.

Party effects

Intra-party conflict

There was little evidence to support H3, which posited that MPs that have higher levels of ideological disagreement with their party and use Twitter as a *substitution channel* have more missing tweets as an accountability avoidance mechanism to avoid sanctioning by their party. While there was a possible association at $p < 0.1$ significance level with ideological distance and missing tweets this effect disappeared when introducing controls. However, general semantic difference was positively associated with missing tweets at the $p < 0.1$ significance level. This may suggest that communicative differences rather than ideological differences may affect the rates of

missing tweets. These results must be interpreted with caution due to their low significance thresholds.

If there was strong evidence to support this hypothesis, it could indicate that MPs attempt to avoid accountability from the party leadership by retroactively removing tweets that may result in future sanctioning. Because there is little or mixed evidence of this happening, this result supports the findings in Castanho Silva and Proksch (2021a), that parties have not yet found a way to control MPs from expressing ideological positions outside of party lines on social media (Ibid, p. 14). This finding, in line with the lack of mass deletion found in H1, suggests that there is little evidence that parliamentarians are actively attempting to hide their tweets to avoid accountability. This implies that social media remains useful as a transparency tool for different principals and that politicians continue to be ‘unleashed’ on social media.

Populist parties

The relationship between populist ideology and missing tweets showed a consistent positive relationship across multiple models. Notably, controlling for party structure did not affect these findings. Thus, while H4a (*MPs in parties with higher levels of populism will have more missing tweets than MPs of other parties*) is confirmed, H4c is rejected as the effect of controlling intra-party democracy was not significant, although introducing it as a control in the minimal model (M2) did decrease the significance of the populism variable. There was also no evidence found for H4b, that effect of populism on the number of missing tweets would be mediated by a radical right-wing ideology. This is a puzzling finding and difficult to explain. The first possible explanation is that the sample included there is too much multicollinearity *nativism* and *populism* and not enough variation making meaningful interaction terms between the two difficult.¹² If this is the case the association between missing tweets and populism may still be driven by RRW populist parties, whose dense Twitter networks are characterized by higher levels of bot activity and xenophobic content lead to higher risks of resharing content that is later removed due to violations of Twitter’s code of conduct. Another possibility is that this association is capturing an unobserved variable of a populist communication *style*. In contrast to a *pro*ri conceptualization of populist political actors used in this study, populism as a communication style, where communicative material is categorized as populist or not and political actors are classified as

¹² A correlation matrix of the populism variables and alternative regression models are provided in the annex.

populist by the intensity of their use of populist communication (de Vreese et al., 2018, p. 526). As the data underlying the *populism* variable comes from an expert survey, where respondents rated parties on a continuous scale, these respondents may have looked to communicative material to inform their answers, especially for ‘grey area’ cases. These results could mean that this has less to do with the ideological underpinnings of the party but the utilization of a specific style by individual MPs. This in combination with the finding in H3 (intra-party conflict) that MPs' general semantic differences from their party on Twitter are associated with missing tweets (significant at the $p < 0.1$ level) suggests that the communicative differences rather than party organization or ideological differences may be an important factor in understanding variations in the number of missing tweets among MPs.

Broader implications

The results of this study can help inform discussions around the broader democratic implications of politicians' removed material on social media. H1, which found no evidence of widespread mass deletion by politicians is most relevant for this discussion. For transparency skeptics, this may be seen as an indication that restricting the right to be forgotten on social media for politicians is unnecessary. Doing so would limit politicians' right to privacy despite there being no widespread attempts at secrecy, with politicians already acting transparently by not mass removing posts. By preserving material that may have little democratic importance, skeptics would argue that this may increase confusion and sow distrust in politicians while harming the privacy rights of deleted retweets by members of the public. On the other hand, there are strong indications that the number of missing tweets is tied to the number of terms one is in office, both in the regression analysis and the finding in H1 that 52.2% ($n=23$) of accounts with no tweets remaining had left politics. This raises normative questions of whether these archives should be preserved, and whether transparency and accountability continue to be democratically important for those who have left parliament. Lastly, it can be argued that the positive association with a populist ideology, which often embraces an illiberal vision of democracy, increases the importance of preserving these materials to guard against democratic threats.

From an empirical standpoint, the results of this study are an important contribution to the discussion of the validity of Twitter as a data source. This study found some evidence that the rate of missing tweets is not randomly distributed. This study showed that there were consistent associations with the number of terms in office and missing tweets and that 52% of the active

Twitter accounts with no tweets remaining had left politics, while populism was positively associated with the number of missing tweets. Not accounting for these non-random removals may bias any dataset using data that was not retrieved in real-time. Therefore, knowledge about these gaps is essential for backward-looking social media research. While previous research has found that the aggregated textual content remains representative (see Zubiaga, 2018), further research is required to know if this textual representativeness generalizes to politicians specifically, who differ in Twitter behavior compared to general users.

Limitations and suggestions for future research

Limitations

There are four major limitations to the study. Firstly, the existing research on the specific theoretical motivations behind social media post removals by politicians is lacking. Because of this many of the hypotheses were derived from Ringel and Davidson's (2020) study of journalists. While some of the motivations are plausibly generalizable to politicians there are still differences in the motivations and communicative logic between politicians and journalists.

The second major limitation of this study is the lack of information about the textual data and metadata of the missing tweets. This results in the dependent variable being a coarse measurement, as it cannot differentiate between active removal and networked effects. Because of the lack of ability to disaggregate between these two mechanisms, there is more noise in the data. Furthermore, because the Twitter ToS requires users to remove locally stored deleted tweets as soon as they are made aware of them being deleted it is impossible to determine even simple descriptive information about the nature of these tweets (i.e., how many are retweets) without violating the Twitter ToS. By examining the number of accounts that practiced mass deletion this study has shed some light on the (lack of) prevalence of mass deletion but more surgical removal of tweets remains unexplored. Because of this, these results underestimate the effects of surgical removals of specific posts.

A third limitation is the temporal span of the study. The tweet removal patterns being examined are at or over a three-year time span. This may introduce more noise, with a higher level of missing tweets from network effects of removed retweets which may obscure shorter interval removals. These short-term removals might be more theoretically relevant for H3-H4, as they may be more relevant as an accountability removal mechanism, than tweets from several years

previously, as immediate tweets are more likely to be held to more scrutiny than all but the most controversial tweets from years past. Jacobs and Spierings (2016) found in their study of Dutch political parties, that social media managers would actively monitor all of the tweets that MPs posted, and when there was discord, actively calling MPs up and telling them to stop tweeting. As social media is fast-moving, party officials who monitor MPs' Twitter activity, are more likely to be responsive to immediate tweets that are out of line with the party's communication strategy and less likely to review older tweets that 'slipped by' the initial party monitoring.

Lastly, case selection was highly restricted due to the data availability. This was particularly limiting for the populism hypotheses as key countries with active left-wing populist parties such as Spain (Podemos), Greece (SYRIZA), and the Netherlands (Socialist Party) were not included in the sample. These parties would add variation to better tease out both the effects of intra-party democracy and nativism in relation to populism, as both Podemos and SYRIZA are highly populist but low in nativism and high in intra-party democracy, while the Dutch Socialist Party has lower levels of intra-party democracy and higher levels of nativism. A broader case selection could confirm the robustness of the findings of H4 (populism).

[Suggestions for future research](#)

There are numerous research avenues that could be taken to examine missing posts on social media. Firstly, this study has shown that there is a lack of knowledge about how politicians view the permanency of their online social media posts. Here, additional qualitative work replicating Ringel and Davidson (2020) but with politicians would greatly contribute to this area of research. This research could shed light on how politicians view the permanency of social media and norms regarding actively removing posts.

Empirically, many of the possible research opportunities are limited due to Twitter's ToS as access to deleted material is essential for enhancing this area of study. To get access to these data, it would be best for future researchers to try and collaborate directly with Twitter to get exceptions to the ToS constraints regarding the handling of deleted material. While some researchers have debated the ethics of breaking ToS agreements, advocating for such an approach disadvantages researchers in more vulnerable professional positions as it increases the risk that API access is shut off by Twitter (Freelon, 2021; Vaccaro et al., 2015). If access is given, researchers could track politicians' Twitter accounts in real-time. As a data generation procedure,

this would provide temporarily fine-grain data in which temporal changes in the metadata (associated responses, likes, and retweets) could provide further parameters to inform the dynamics of tweet deletion.

With this higher quality data, many of the hypotheses examined here could be retested using more granular data. H3 (intra-party conflict) could be tested further as the attitudinal dimensions of actively deleted tweets specifically by individual MPs could be compared at scale to the overall party. If for instance, the content of the deleted tweets is ideologically farther removed from the MPs party than non-deleted material, this may indicate some level of more surgical tweet removals. For H2 (gendered incivility) these data with a temporal dimension could be used to see if exposure to incivil tweets later leads to more active post removals. Lastly, for H4 (populism) access to the textual data of deleted content would allow researchers to examine populism as a *communication style*, allowing researchers to code tweets according to different dimensions of populism at the MP level (see Ernst et al., 2017). This would allow levels of populism to be calculated at the MP level, which could measure intra-party variation in the amount of populist rhetoric deployed, which could be a fruitful ground for research on populism that moves beyond the *a priori* conceptualization of populist parties that was used in this study.

Additionally, this study also raises broader questions on ephemerality on social media and how the rights of privacy and transparency should be balanced for public officials on social media. There has been much written about platform transparency regarding electoral ad spending and microtargeting. However, how policies guarding the “right to disappear” to politicians and the tensions the tension that this raises between transparency and privacy have not been widely analyzed. While this tension was discussed in the *Theoretical background* of this paper, more research is needed to understand the regulatory dynamics between platforms ToS, data protection laws, and transparency as well as the tensions between privacy rights and transparency. Due to the association with missing tweets and leaving parliament, future research could also explore the democratic implications of this behavior, especially in the context of the ‘revolving-door’ relationship between private business, lobbying, and post-parliamentary careers (Rasmussen et al., 2021, p. 488).

Lastly, outside of Twitter, many platforms such as Snapchat are intrinsically ephemeral content and older platforms are introducing these features. The users of these platforms tend to skew

young and have not yet been widely adopted by politicians. Intrinsicly, this hyper-ephemeral content makes questions about how to balance transparency and privacy in social media more salient. If social media with historical records “unleashed” politicians, ephemerally designed platforms may further breakdown intensify the ability for individual MPs to express themselves against the party lines. By totally removing any records of what was said, these platforms make it more difficult for party social media managers to police, who cannot retroactive review how an MP conducts themselves on these ephemeral platforms. These platforms should not be ignored due to bias towards Twitter and Facebook, which are older social media platforms that, due to their mainly textual nature and ease of API access are often more appealing for researchers.

Conclusions

This study is first identified to conduct a systematic multi-country analysis examining the factors behind politicians’ missing tweets. The research question’s first point of inquiry was the extent of missing tweets among parliamentarians in Western Europe. This was answered after the data collection process which found that 21.8% of the sample’s tweets from 2018 had disappeared from the platform by October 2021, which is roughly in line with other estimates of dataset missingness. This finding is somewhat surprising as other studies of missing tweets focus on the general Twitter userbase, introducing issues with spam, bot accounts, and other activity that may promote post removal. Empirically this suggests that future research on parliamentarians that uses old Twitter data must be mindful of how this missingness affects the validity of their data. While this study determined that politicians do have missing tweets, the extent of active deletion of posts is unclear. Notably, contrary to expectation, parliamentarians in Western Europe do not widely engage in mass deletion of their tweets. While this does not preclude more surgical removals of specific posts, it does show that there is not a norm of mass deletion of tweets in any subgroup of the sample studied.

In answering what factors lay behind these removals, this study found that there was no significant relationship found between the number of missing tweets and gender or intra-party conflict. While there was a statistically significant positive association with populism and the number of missing tweets, this could not be explained by party structure and was not conditional on nativist ideology. This may be due to methodological issues due to high levels of multicollinearity *nativism* and *populism*. If this is the case, the association between populism and

missing tweets may still be driven by radical right-wing populist parties, whose dense twitter networks are rife with xenophobia and bot activity. Alternatively, this may be due to the measurement of populism as a continuous scale as opposed to a dichotomous measurement indirectly capturing populism as a communicative *style*. This together with the finding that MPs semantic differences from their party were positively associated with post removals suggests that future research could benefit from analyzing the relationship between populism and post-removal using a communication-centered approach to populism at the MP level.

The results of this study can help inform the broader discussion of the democratic implications of digital privacy rights, which protect politicians deleted material from being archived, and transparency. This study found that mass deletion of tweets was not common among MPs, they are not using this blunt tool to remove their social media records. Thus, the justification for a reduction of politicians' privacy rights is lower while increasing the probability that preserving their tweets may include interactions with users who are not public officials as well as innocuous material that has little democratic utility. However, the association between leaving parliament and mass deletion, as well as the positive association with a populist ideology, suggests that there may be democratic utility in preserving these tweets. Furthermore, the limitations of the Twitter ToS and data protection laws actively limit further examination needed to determine if this material is in the public interest to preserve.

Taken as a whole, this study reaffirms the possibility of social media as a transparency and accountability tool, allowing voters, journalists, and party elites insight into the opinions and attitudes of parliamentarians both in real-time and in the past. For party leaders and elites, this finding suggests that the lack of mass deletion or positive association with ideological divergence and removal patterns, suggests that politicians are not worried about expressing dissenting opinions online. For voters, access to these dissenting opinions can shine a light on the factions of the parties that they vote for.

While future research is required to confirm these findings in different settings and under different timespans with more disaggregated data, this thesis has shown that the study of missing tweets is an important field of research for both understanding politician's social media behavior as well as the wider democratic consequences of increasingly digitized political communication and its ephemerality.

Bibliography

- Ahmed, R., Pisiu, D., 2021. Uniting the far right: how the far-right extremist, New Right, and populist frames overlap on Twitter – a German case study. *European Societies* 23, 232–254. <https://doi.org/10.1080/14616696.2020.1818112>
- Åkerlund, M., 2020. The importance of influential users in (re)producing Swedish far-right discourse on Twitter. *European Journal of Communication* 35, 613–628. <https://doi.org/10.1177/0267323120940909>
- Almuhimedi, H., Wilson, S., Liu, B., Sadeh, N., Acquisti, A., 2013. Tweets are forever: a large-scale quantitative analysis of deleted tweets, in: *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW '13*. Presented at the the 2013 conference, ACM Press, San Antonio, Texas, USA, p. 897. <https://doi.org/10.1145/2441776.2441878>
- Ananny, M., Crawford, K., 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20, 973–989. <https://doi.org/10.1177/1461444816676645>
- Barberá, P., 2015. Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Polit. anal.* 23, 76–91. <https://doi.org/10.1093/pan/mpu011>
- Barberá, P., Rivero, G., 2015. Understanding the Political Representativeness of Twitter Users. *Social Science Computer Review* 33, 712–729. <https://doi.org/10.1177/0894439314558836>
- Barberá, P., Steinert-Threlkeld, Z.C., 2020. How to Use Social Media Data for Political Science Research, in: *The SAGE Handbook of Research Methods in Political Science and International Relations*. SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road London EC1Y 1SP, pp. 404–423. <https://doi.org/10.4135/9781526486387.n26>
- Barberá, P., Zeitzoff, T., 2018. The New Public Address System: Why Do World Leaders Adopt Social Media? *International Studies Quarterly* 62, 121–130. <https://doi.org/10.1093/isq/sqx047>
- Bar-On, T., 2018. *The Radical Right and Nationalism*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190274559.013.2>
- Baume, S., Papadopoulos, Y., 2018. Transparency: from Bentham's inventory of virtuous effects to contemporary evidence-based scepticism. *Critical Review of International Social and Political Philosophy* 21, 169–192. <https://doi.org/10.1080/13698230.2015.1092656>
- Békés, G., Kézdi, G., 2021. *Data analysis for business, economics, and policy*. Cambridge University Press, Cambridge, United Kingdom ; New York, NY.
- Blair, G., Cooper, J., Coppock, A., Humphreys, M., Sonnet, L., 2021. *estimatr: Fast estimators for design-based inference (manual)*.
- Bleakley, P., 2021. Panic, pizza and mainstreaming the alt-right: A social media analysis of Pizzagate and the rise of the QAnon conspiracy. *Current Sociology* 001139212110348. <https://doi.org/10.1177/00113921211034896>
- Bossetta, M., 2018. The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election. *Journalism & Mass Communication Quarterly* 95, 471–496. <https://doi.org/10.1177/1077699018763307>
- Carey, J.M., 2008. *Legislative Voting and Accountability*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511810077>

- Castanho Silva, B., Proksch, S.-O., 2021a. Politicians unleashed? Political communication on Twitter and in parliament in Western Europe. *PSRM* 1–17. <https://doi.org/10.1017/psrm.2021.36>
- Castanho Silva, B., Proksch, S.-O., 2021b. Fake It ‘Til You Make It: A Natural Experiment to Identify European Politicians’ Benefit from Twitter Bots. *Am Polit Sci Rev* 115, 316–322. <https://doi.org/10.1017/S0003055420000817>
- Ceron, A., 2017a. *Social Media and Political Accountability*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-52627-0>
- Ceron, A., 2017b. Intra-party politics in 140 characters. *Party Politics* 23, 7–17. <https://doi.org/10.1177/1354068816654325>
- Chang, W., 2019. *webshot: Take screenshots of web pages (manual)*.
- Crowell, C., 2015. Holding public officials accountable with Twitter and Politwoops [WWW Document]. URL https://blog.twitter.com/en_us/a/2015/holding-public-officials-accountable-with-twitter-and-politwoops (accessed 10.6.21).
- de Vreese, C.H., Esser, F., Aalberg, T., Reinemann, C., Stanyer, J., 2018. Populism as an Expression of Political Communication Content and Style: A New Perspective. *The International Journal of Press/Politics* 23, 423–438. <https://doi.org/10.1177/1940161218790035>
- Döring, H., Manow, P., 2018. ParlGov 2018 Release. <https://doi.org/10.7910/DVN/F0YGNC>
- Engesser, S., Ernst, N., Esser, F., Büchel, F., 2017. Populism and social media: how politicians spread a fragmented ideology. *Information, Communication & Society* 20, 1109–1126. <https://doi.org/10.1080/1369118X.2016.1207697>
- Enli, G.S., Skogerbø, E., 2013. Personalized Campaigns in Party-centred Politics: Twitter and Facebook as arenas for political communication. *Information, Communication & Society* 16, 757–774. <https://doi.org/10.1080/1369118X.2013.782330>
- Ernst, N., Engesser, S., Büchel, F., Blassnig, S., Esser, F., 2017. Extreme parties and populism: an analysis of Facebook and Twitter across six countries. *Information, Communication & Society* 20, 1347–1364. <https://doi.org/10.1080/1369118X.2017.1329333>
- Etzioni, A., 2010. Is Transparency the Best Disinfectant? *Journal of Political Philosophy* 18, 389–404. <https://doi.org/10.1111/j.1467-9760.2010.00366.x>
- European Commission, 2016. *General Data Protection Regulation (GDPR): Regulation (EU) 2016/679*.
- Fearon, J.D., 1999. Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance, in: Przeworski, A., Stokes, S.C., Manin, B. (Eds.), *Democracy, Accountability, and Representation*. Cambridge University Press, pp. 55–97. <https://doi.org/10.1017/CBO9781139175104.003>
- Forgey, Q., 2021. National Archives can’t resurrect Trump’s tweets, Twitter says [WWW Document]. *POLITICO*. URL <https://www.politico.com/news/2021/04/07/twitter-national-archives-realdonaldtrump-479743> (accessed 12.27.21).
- Freelon, D., 2021. When Social Media Data Disappear, in: *Research Exposed*. Columbia University Press, pp. 6–29. <https://doi.org/10.7312/harg18876-002>
- Gailmard, S., 2014. Accountability and Principal–Agent Theory, in: Bovens, M., Goodin, R.E., Schillemans, T. (Eds.), *The Oxford Handbook of Public Accountability*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199641253.013.0016>
- Golbeck, J., Grimes, J.M., Rogers, A., 2010. Twitter use by the U.S. Congress. *J. Am. Soc. Inf. Sci.* n/a-n/a. <https://doi.org/10.1002/asi.21344>

- Gorwa, R., Ash, T.G., 2020. Democratic Transparency in the Platform Society, in: Tucker, J.A., Persily, N. (Eds.), *Social Media and Democracy: The State of the Field, Prospects for Reform*, SSRC *Anxieties of Democracy*. Cambridge University Press, Cambridge, pp. 286–312.
- Heiss, R., Matthes, J., 2020. Stuck in a Nativist Spiral: Content, Selection, and Effects of Right-Wing Populists' Communication on Facebook. *Political Communication* 37, 303–328. <https://doi.org/10.1080/10584609.2019.1661890>
- Hern, A., 2015. Twitter blocks access to political transparency organisation Politwoops. *The Guardian*.
- Jacobs, K., Sandberg, L., Spierings, N., 2020. Twitter and Facebook: Populists' double-barreled gun? *New Media & Society* 22, 611–633. <https://doi.org/10.1177/1461444819893991>
- Jacobs, K., Spierings, N., 2019. A populist paradise? Examining populists' Twitter adoption and use. *Information, Communication & Society* 22, 1681–1696. <https://doi.org/10.1080/1369118X.2018.1449883>
- Jacobs, K., Spierings, N., 2016. *Social media, parties, and political inequalities*. Palgrave Macmillan, Basingstoke New York.
- Jungherr, A., 2016. Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics* 13, 72–91. <https://doi.org/10.1080/19331681.2015.1132401>
- Kearney, M., 2019. rtweet: Collecting and analyzing Twitter data. *JOSS* 4, 1829. <https://doi.org/10.21105/joss.01829>
- Keller, T.R., Klinger, U., 2019. Social Bots in Election Campaigns: Theoretical, Empirical, and Methodological Implications. *Political Communication* 36, 171–189. <https://doi.org/10.1080/10584609.2018.1526238>
- Klinger, U., Svensson, J., 2015. The emergence of network media logic in political communication: A theoretical approach. *New Media & Society* 17, 1241–1257. <https://doi.org/10.1177/1461444814522952>
- Leary, M.R., Kowalski, R.M., 1990. Impression management: A literature review and two-component model. *Psychological Bulletin* 107, 34–47. <https://doi.org/10.1037/0033-2909.107.1.34>
- Lijphart, A., 2012. *Patterns of democracy: government forms and performance in thirty-six countries*, 2nd ed. ed. Yale University Press, New Haven.
- Liu, Y., Kliman-Silver, C., Mislove, A., 2014. The Tweets They Are a-Changin': Evolution of Twitter Users and Behavior. *ICWSM* 8, 305–314.
- Maddock, J., Starbird, K., Mason, R., 2015. Using Historical Twitter Data for Research: Ethical Challenges of Tweet Deletions. Presented at the CSCW '15 Workshop on Ethics at the 2015 Conference on Computer Supported Cooperative Work (CSCW 2015), Vancouver, BC, Canada, p. 6.
- McCammon, M., 2020. Tweeted, deleted: An exploratory study of the US government's digital memory holes. *New Media & Society* 146144482093403. <https://doi.org/10.1177/1461444820934034>
- Meeks, L., 2018. Tweeted, deleted: theoretical, methodological, and ethical considerations for examining politicians' deleted tweets. *Information, Communication & Society* 21, 1–13. <https://doi.org/10.1080/1369118X.2016.1257041>
- Meijer, A., 2014. Transparency, in: Bovens, M., Goodwin, R., Schillemans, T. (Eds.), *The Oxford Handbook of Public Accountability*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199641253.013.0043>

- Meijers, M., Zaslove, A., 2020. Populism and Political Parties Expert Survey 2018 (POPPA). <https://doi.org/10.7910/DVN/8NEL7B>
- Meijers, M.J., Zaslove, A., 2021. Measuring Populism in Political Parties: Appraisal of a New Approach. *Comparative Political Studies* 54, 372–407. <https://doi.org/10.1177/0010414020938081>
- Mellon, J., Prosser, C., 2017. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics* 4, 205316801772000. <https://doi.org/10.1177/2053168017720008>
- Metz, M., Kruike-meier, S., Lecheler, S., 2020. Personalization of politics on Facebook: examining the content and effects of professional, emotional and private self-personalization. *Information, Communication & Society* 23, 1481–1498. <https://doi.org/10.1080/1369118X.2019.1581244>
- Mikal, J., Hurst, S., Conway, M., 2016. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. *BMC Med Ethics* 17, 22. <https://doi.org/10.1186/s12910-016-0105-5>
- Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., Rosenquist, J., 2021. Understanding the Demographics of Twitter Users. *ICWSM* 5, 554–557.
- Mudde, C., 2017. *Populism*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198803560.013.1>
- Nyman, P., 2021. More about regression.
- Paul, K., 2021. National Archives won't be allowed to restore Trump's tweets on the platform. *The Guardian*.
- Politwoops, n.d. *Politwoops: FAQ [WWW Document]*. Politwoops. URL <https://politwoops.eu/page/faq> (accessed 12.27.21).
- Rasmussen, A., Buhmann-Holmes, N., Egerod, B.C.K., 2021. The Executive Revolving Door: New Dataset on the Career Moves of Former Danish Ministers and Permanent Secretaries. *Scandinavian Political Studies* 44, 487–502. <https://doi.org/10.1111/1467-9477.12214>
- Reventlow, N.J., 2020. Can the GDPR and Freedom of Expression Coexist? *AJIL Unbound* 114, 31–34. <https://doi.org/10.1017/aju.2019.77>
- Ringel, S., Davidson, R., 2020. Proactive ephemerality: How journalists use automated and manual tweet deletion to minimize risk and its consequences for social media as a public archive. *New Media & Society* 146144482097238. <https://doi.org/10.1177/1461444820972389>
- Rydgren, J., 2007. The Sociology of the Radical Right. *Annu. Rev. Sociol.* 33, 241–262. <https://doi.org/10.1146/annurev.soc.33.040406.131752>
- Sältzer, M., 2020. Finding the bird's wings: Dimensions of factional conflict on Twitter. *Party Politics* 135406882095796. <https://doi.org/10.1177/1354068820957960>
- Schudson, M., 2020. The Shortcomings of Transparency for Democracy. *American Behavioral Scientist* 64, 1670–1678. <https://doi.org/10.1177/0002764220945347>
- Skaaning, S.-E., 2021. Democracy: contested concept with a common core, in: *Research Handbook on Democracy and Development*. Edward Elgar Publishing, pp. 27–44. <https://doi.org/10.4337/9781788112659.00008>
- Southern, R., Harmer, E., 2021. Twitter, Incivility and “Everyday” Gendered Othering: An Analysis of Tweets Sent to UK Members of Parliament. *Social Science Computer Review* 39, 259–275. <https://doi.org/10.1177/0894439319865519>

- Stanyer, J., Salgado, S., Strömbäck, J., 2016. Populist Actors as Communicators or Political Actors as Populist Communicators: Cross-National Findings and Perspectives, in: *Populist Political Communication in Europe*. Routledge.
<https://doi.org/10.4324/9781315623016>
- Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S.A., 2020. The Dynamics of Political Incivility on Twitter. *SAGE Open* 10, 215824402091944.
<https://doi.org/10.1177/2158244020919447>
- Thompson, J.B., 2011. Shifting Boundaries of Public and Private Life. *Theory, Culture & Society* 28, 49–70. <https://doi.org/10.1177/0263276411408446>
- Timoneda, J.C., 2018. Where in the world is my tweet: Detecting irregular removal patterns on Twitter. *PLoS ONE* 13, e0203104. <https://doi.org/10.1371/journal.pone.0203104>
- Tromble, R., 2018a. Thanks for (actually) responding! How citizen demand shapes politicians' interactive practices on Twitter. *New Media & Society* 20, 676–697.
<https://doi.org/10.1177/1461444816669158>
- Tromble, R., 2018b. The great leveler? Comparing citizen–politician Twitter engagement across three Western democracies. *Eur Polit Sci* 17, 223–239. <https://doi.org/10.1057/s41304-016-0022-6>
- Tucker, J.A., Theocharis, Y., Roberts, M.E., Barberá, P., 2017. From Liberation to Turmoil: Social Media And Democracy. *Journal of Democracy* 28, 46–59.
<https://doi.org/10.1353/jod.2017.0064>
- Twitter, 2021a. About election labels on Twitter [WWW Document]. Twitter. URL <https://help.twitter.com/en/using-twitter/election-labels> (accessed 12.23.21).
- Twitter, 2021b. Government and state-affiliated media account labels [WWW Document]. Twitter. URL <https://help.twitter.com/en/rules-and-policies/state-affiliated> (accessed 12.23.21).
- Twitter, 2020. Developer Agreement and Policy – Twitter Developers [WWW Document]. URL <https://developer.twitter.com/en/developer-terms/agreement-and-policy> (accessed 10.6.21).
- Uršič, H., 2016. The Right to be Forgotten or the Duty to be Remembered? Twitter data reuse and implications for user privacy.
- Vaccaro, K., Karahalios, K., Sundvig, C., Hamilton, K., Langbort, C., 2015. Agree or cancel? Research and terms of service compliance. Presented at the CSCW 2015 Workshop: Ethics for Studying Sociotechnical Systems in a Big Data World, Vancouver, BC, Canada.
- Ward, S., McLoughlin, L., 2020. Turds, traitors and tossers: the abuse of UK MPs via Twitter. *The Journal of Legislative Studies* 26, 47–73.
<https://doi.org/10.1080/13572334.2020.1730502>
- Weller, K. (Ed.), 2013. *Twitter and society, Digital formations*. Peter Lang, New York.
- Wojcik, S., Hughes, A., 2019. How Twitter Users Compare to the General Public. Pew Research Center: Internet, Science & Tech. URL <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/> (accessed 10.7.21).
- Xie, S., 2020. New conversation settings, coming to a Tweet near you [WWW Document]. URL https://blog.twitter.com/en_us/topics/product/2020/new-conversation-settings-coming-to-a-tweet-near-you (accessed 11.24.21).

Zubiaga, A., 2018. A longitudinal assessment of the persistence of twitter datasets. *Journal of the Association for Information Science and Technology* 69, 974–984.
<https://doi.org/10.1002/asi.24026>

Appendix

A.1. Detailed technical explanation of data gathering and validation procedure

A.1.1. Data identification and gathering

This section outlines the data gathering process with more technical depth. Due to Twitter's ToS full tweet datasets cannot be shared between researchers. Only unique Tweet IDs which then are 'rehydrated' or redownloaded can be shared. To rehydrate the Castanho Silva and Proksch (2021a), first, all the RData files provided by Castanho Silva and Proksch (2021a) were loaded into R. These objects are lists, specifically lists used by the `quanteda` package to create document term matrices (DTM) for text analysis. The element of interest in these lists is the ``ids`` element, which consists of a dataframe that has three columns, a tweet ID, the screen name associated with the tweet, and a flag for if the tweet was about the EU or not. The two variables that were of interest for this research are the ID and the screenname. These were extracted to create a new list of just the relevant dataframes. The next step was to collapse this list of dataframes into a single dataframe with a new variable country variable.

One limitation with the Twitter API is that certain functions are rate-limited, meaning there is a limit to how many times a given action can be done in a given time frame. Using the version of the API that the `rtweet` package accesses, users are limited to downloading 90,000 tweets every fifteen minutes (Kearney, 2019). To bypass this limitation, the following steps were taken. First, a new variable was added to the dataset which groups tweet into batches of 90,000 which for the whole dataset totals 24 groups. Next, a function called `batch_downloader` was written. This function takes a dataframe as the argument, and simply wraps `lookup_statuses` from the `rtweet` project in a timer, so that after the `lookup_statuses` function is run, R waits until the rate limit timer expires before running the next operation. The code for this function is available below.


```

batch_downloader <- function(df){
  # Grab the time and print the time
  time_start <- Sys.time()
  print(time_start)

  # Create a list of tweet IDs
  ids <- df$status_id

  # Pull the tweets from Twitter
  tweets <- lookup_statuses(ids, parse = TRUE)

  # Grab the end time and print the end time
  time_end <- Sys.time()
  print(time_end)

  # Calculate the total time the operation has taken
  time_taken <- time_end - time_start
  print(time_taken)

  # Grab the rate limit time add an extra minute for safety
  rl <- rate_limit(query = "lookup_statuses")
  mins_until_rest <- rl$reset
  seconds_rest <- as.numeric(mins_until_rest)*60+5
  print(paste("Sleeping for",round(as.numeric(mins_until_rest),2), "minutes")
)

  # Sleep
  Sys.sleep(seconds_rest)
  return(tweets)}

```

However, this function does not divide the tweets into blocks of 90,000 IDs. Using this function on the full dataset would not work as it would run the whole operation, fail due to rate limiting, and then it would sleep for 15 minutes. To remedy this, the dataset is filtered by country and then reconverted into a list using the *group_split* function from the *dplyr* package. The result is a list consisting of elements of dataframes, each of which contains less than 90,000 tweets. Next, the *map* function from the *purrr* package is used to apply the *batch_downloader* to each element (the batched dataframes) in the list. The code below provides an example of this procedure.

```

# Sweden Tweets 2018 n = 132,737
sweden_ids_df <- filter(country_tweet_ids_df, country.r == "sweden")

# Group into batches <90,000
df_group_sweden <- group_split(sweden_ids_df, id_group_num)

# Iterate through list
sweden_tweets_list <- map(df_group_sweden, batch_downloader)

```

This procedure is done for all countries. It should be noted that many inefficiencies in the current code could be streamlined if needed. It currently takes over seven hours in total to download all of the tweets from the seven countries in the original sample. This code has been built to ensure that the rate limit is not reached rather than for efficiency. As this dataset only includes 2 million tweets these inefficiencies are manageable, however, larger-scale studies should look for alternative solutions.

Once all country tweet datasets are produced, a new dataset is created with just the counts of existing tweets per parliamentarian in the 2021 sample. This is then merged with the aggregate dataset from Castanho Silva and Proksch (2021a) which provides parliamentarian level data on the number of tweets in 2018 along with relevant demographic and party information along with aggregate twitter sentiment scores calculated for their research.

A.1.2. Cleaning and validation

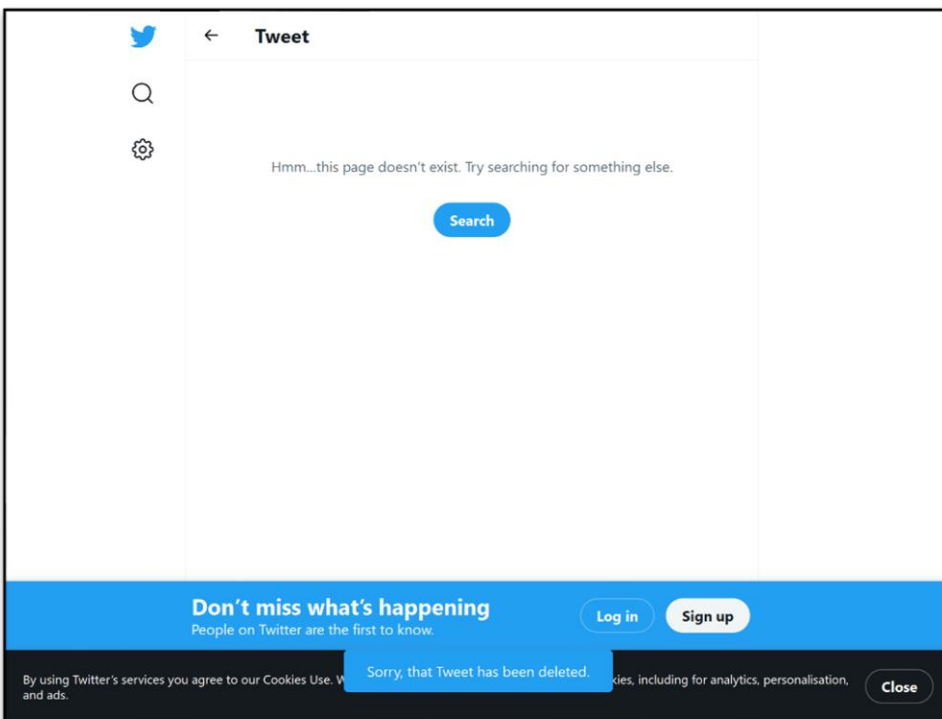
To ensure that there were no coding errors at the initial stages of data gathering, the variable with the counts of tweets per parliamentarian in the aggregate dataset from Castanho Silva and Proksch (2021a) was compared to counts per parliamentarians that were generated from the RData files from which the ID statuses were sourced. The results of this showed that there was no deviation between the two sets of counts for the 2018 amount of tweets.

One complication is that Twitter allows users to change their screen names. However, the unique status IDs remain the same, and when using the *lookup_statuses* function tweets are still downloaded, but now are associated with a new screenname. As an illustrative example Sir Nicholas Dakin, a Labour politician from the UK, changed his screenname from *nicdakinmp* to *nicdakin55* sometime between 2018 and 2021. Screenname changes were identified by looking at accounts that had more than 2 tweets in the 2021 dataset but did not match any screennames in the 2018 dataset. The new screen names were identified from simple internet searches, starting with searching for the old screen name, followed by searching for the name of the politician on

Twitter. In total 158 users' screen name changes were identified. Once identified these screennames were added to the dataset in a new variable that added the 2018 screen names to the cases where it changed. One limitation of this validation approach is that it can only identify users who changed their screennames using the official Twitter function. It is possible that some politicians created new accounts with new screennames.

The next validation step was to validate that the tweets were missing. The Twitter API, at least accessed through the `rtweet` package, does not provide a flag when a given Tweet ID does not return any data. To validate that these tweets were missing, a random sample of 1,000 tweets was drawn from tweets that were flagged as removed. To try and expedite this validation process, it was first attempted to create a web scraper that would scrape the content of each of the tweet pages to determine if these tweets were missing. However, it became quickly apparent that this was a very difficult process due to the use of dynamically changing HTML tags which prevents simple web scraping. Instead, the `webshots` package for R was used in combination with a for loop to take screenshots of all of the sampled tweets. The results of this were then hand-coded to ensure that each of the tweets in the sample was missing

A webshots screen shot of a missing tweet



There were also cases where the tweet ID was reused for another account. This did not affect the data on missing tweets because the counts were generated by counting the number of posts per screenname. The screenshot below shows a case of a reused Tweet ID. This is a Tweet ID that in the 2018 dataset was associated with Jean-Luc Mélenchon, the leader of the French far-left party La France Insoumise. Here, it is clear that this is not the original tweet, however, during the manual coding of these tweets, any ambiguity at all warranted further investigation which usually was done by checking the account associated with the reused count to ensure it was not the MP in question.

A webshots screenshot of a tweet with a reused ID



A version of this validation procedure was used in the initial data collection and before being systematized during the final data collection. The data was first collected in late September 2021. The shorter version of these validation checks revealed that there was an issue with the code, with many tweets being falsely coded as missing in the Danish sample. This was due to an error in the code in which the tweet IDs were read as numeric class variables and not character class variables. As there is a temporal dimension to the dependent variable, this mistake required that all of the datasets were redownloaded again around the same time. Following this, the validation

checks were repeated systematically. It was during this validation period when was seen that the account associated with Boris Johnson was the account of the Foreign Office and that there were issues with duplication of the accounts of French MP Valérie Oppelt and German MP Michael Grosse-Brömer.

A.2. Tweet Inequality

Gini coefficient by country

Country	Gini coefficient tweets 2018
Denmark	0.621036
France	0.494855
Germany	0.695211
Italy	0.677258
Sweden	0.713447
UK	0.500496

A.3. Mass Deletion

There was some country level variation in the number of MPs with *no remaining tweets accessible*. The table below shows these differences. Germany and the United Kingdom had an overrepresentation of MPs removing all their tweets, while France and Italy had an underrepresentation of tweet removal.

Country	Total MPs with over 100 tweets	MPs with over 100 tweets with none remaining in 2021	% Difference
Germany	16.2% (267)	29.5% (13)	13.40%
UK	28.6% (472)	36.4% (16)	7.80%
Sweden	8.7% (144)	11.4% (5)	2.60%
Denmark	5.5% (91)	4.5% (2)	-1%
Italy	14.3% (237)	4.5% (2)	-9.80%
France	26.7% (441)	13.6% (6)	-13.10%

A.4. Regression analysis

A.4.1. Description of variables and 2018 regression models

Overview of variables used in regression models

Variable	N = 2,251¹
N missing tweets in 2021 (log)	4.43 (1.79)
Male	1,448 / 2,209 (66%)
(Missing)	42
N tweets in 2018 (log)	5.95 (1.84)
Minister (ever)	312 / 2,203 (14%)
(Missing)	48
Leadership position	51 / 2,251 (2.3%)
Terms in parliament	2.00 (1.58)
(Missing)	48
Party in government	1,129 / 2,157 (52%)
(Missing)	94
Share of seats held by party	0.20 (0.14)
(Missing)	89
Party ideology left to right (scaled)	0.22 (0.87)
(Missing)	97
EU Dissent on Twitter	0.34 (0.48)
(Missing)	290
Distance to party on Twitter (non-EU)	0.23 (0.30)
Populism (scaled)	-0.21 (0.94)
(Missing)	97
Intra-party democracy	-0.38 (0.92)
(Missing)	101
Nativism (scaled)	-0.29 (0.83)
(Missing)	97

¹Median (SD); n / N (%)

Regression models for total tweets posted in 2018

DV:# Tweets in 2018 (log)								
Variable	M1: MP		M2: Party		M3: Populism		M4: Populism and Intra-party Democracy	
	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²
Male	-0.25**	0.108	-0.11	0.085	-0.09	0.076	-0.09	0.078
Minister (ever)	0.41***	0.149	0.49***	0.146	0.47***	0.142	0.47***	0.142
Leadership position	1.1****	0.159	1.1****	0.162	1.1****	0.161	1.1****	0.161
Terms in parliament	-0.04	0.050	-0.05	0.049	-0.06	0.047	-0.06	0.046
Party in government			-0.36*	0.183	-0.40**	0.162	-0.40**	0.163
Share of seats held by party			0.47	0.488	0.46	0.501	0.41	0.575
Party ideology left to right (scaled)			-0.29***	0.094	-0.26***	0.077	-0.28***	0.086
Populism (scaled)					-0.15*	0.083	-0.16*	0.093
Intra-party democracy (scaled)							-0.02	0.107
Adjusted R ²	0.204		0.232		0.236		0.236	

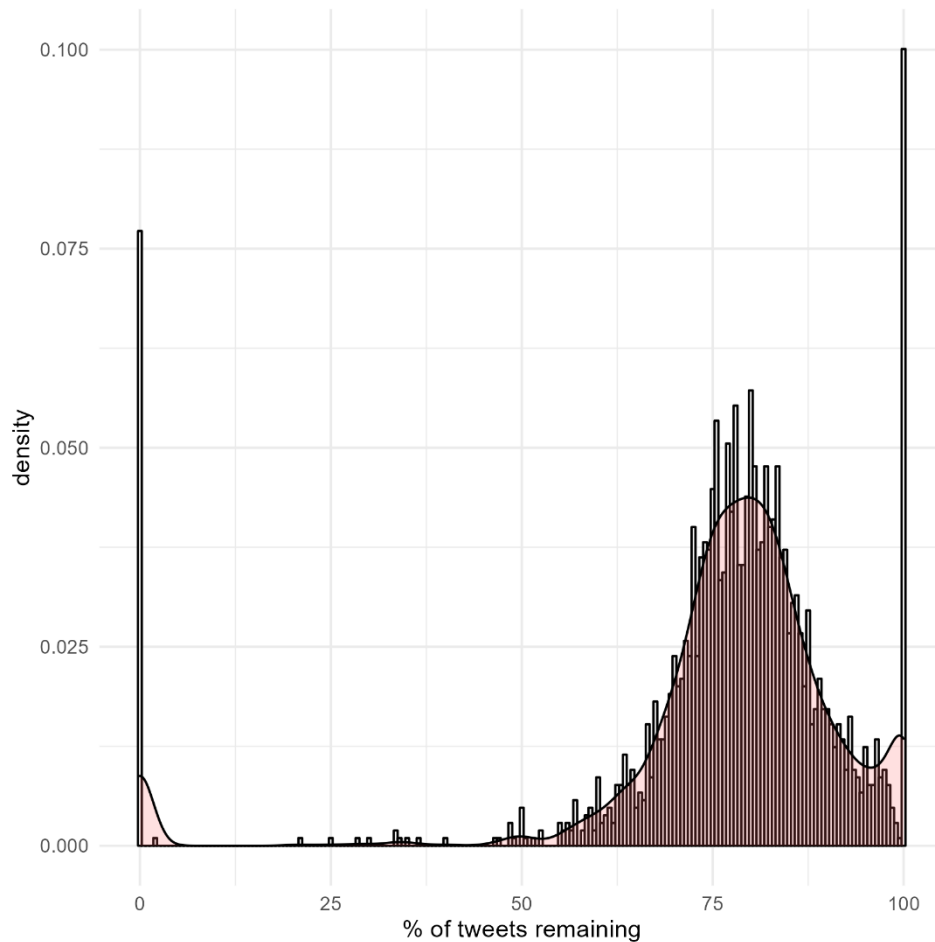
¹*p<0.1; **p<0.05; ***p<0.01; ****p<0.001

²SE = Standard Error

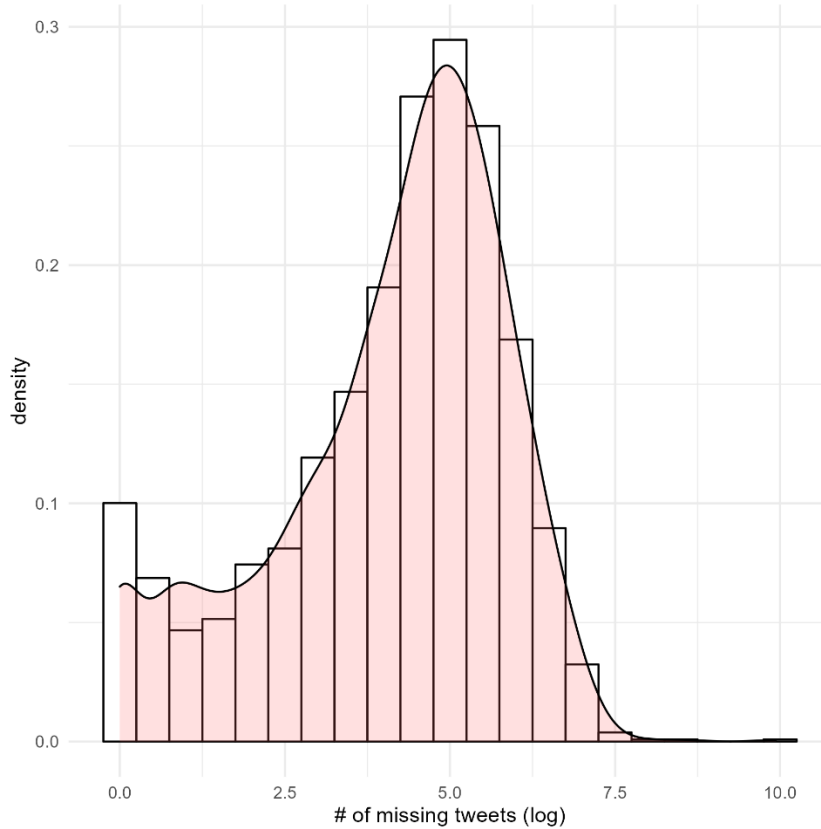
A.4.2. Regression models with an alternative dependent variable

This section presents the regression models using the percentage of tweets remaining as the dependent variable. The figures below present the distributions of the two variables. The distribution of the percentage of tweets remaining has high peaks on both ends of the distribution. In contrast, the logged number of missing tweets has a much smoother distribution, although it has a heavy right tail.

Distribution of % of tweets remaining variable



Distribution of # of missing tweets (log) variable



The results of the regressions with the alternate dependent variable specification are presented below. As the dependent variable is the percentage of tweets remaining, positive coefficients indicate less missingness and negative coefficients indicate more. Between the two specifications, the top-line results remain consistent, with populism the only hypothesized variable of interest to reach $p < .05$ significance. Additionally, the association between missingness and the number of terms served remains robust. However, the alternative specification indicates higher levels of significance for the leadership position variable, which remains significant across a range of models and is positively associated with lower levels of missingness.

H2: Gender, alternative DV

DV: % of tweets remaining						
Variable	M1: Gender		M2: MP controls		M3: MP and party controls	
	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²
Male	-0.79	0.890	-0.64	0.867	-0.63	0.905
N tweets in 2018 (log)	0.50	0.514	0.45	0.510	0.42	0.526
Minister (ever)			1.1	1.04	1.4	1.08
Leadership position			3.2**	1.26	3.0**	1.29
Terms in parliament			-0.54	0.328	-0.59*	0.321
Party in government					-2.7*	1.54
Share of seats held by party					0.54	4.48
Party ideology left to right (scaled)					0.74	0.763
N	2,098		2,098		2,098	
Adjusted R ²	0.869		0.869		0.869	

¹*p<0.1; **p<0.05; ***p<0.01; ****p<0.001

²SE = Standard Error

H3: Intra-party conflict, alternative DV

Variable	DV: % of tweets remaining					
	M1: Difference to Party (EU)		M2: Difference to Party (EU) Controls		M3: Diff to Party (Tone Controlled)	
	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²
EU Dissent on Twitter	-0.92	0.98	-0.70	0.98	-0.47	0.904
Distance to party on Twitter (non-EU)					-3.0	2.61
N tweets in 2018 (log)	-0.03	0.539	-0.16	0.541	-0.24	0.597
Terms in parliament			-0.56*	0.287	-0.53*	0.282
Male			-0.63	0.787	-0.56	0.788
Minister (ever)			1.0	1.15	1.1	1.15
Leadership position			3.3**	1.28	3.2**	1.25
Party in government			-0.68	1.45	-0.53	1.41
Share of seats held by party			-0.15	4.39	-0.51	4.42
Party ideology left to right (scaled)			-0.45	0.712	-0.40	0.706
N	1,862		1,862		1,862	
Adjusted R ²	0.004		0.005		0.006	

¹*p<0.1; **p<0.05; ***p<0.01; ****p<0.001

²SE = Standard Error

H4: Populism, alternative DV

DV: % of tweets remaining										
Variable	M1: Populism		M2: Populism and intra-party democracy		M3: Populism with controls		M4: Populism and intra-party democracy with controls		M5: Populism x nativism Interaction	
	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²
Populism (scaled)	-1.2*	0.600	-1.2*	0.664	-1.6**	0.641	-1.5**	0.739	0.81	1.81
Intra-party democracy (scaled)			0.14	0.606			0.36	0.902		
Populism X Nativism									-3.2	2.50
Nativism (scaled)									0.30	1.58
N tweets in 2018 (log)	0.49	0.541	0.48	0.545	0.38	0.538	0.38	0.539	0.37	0.537
Terms in parliament					-0.68**	0.318	-0.68**	0.313	-0.66**	0.311
Male					-0.30	0.916	-0.28	0.910	-0.32	0.912
Minister (ever)					1.2	1.07	1.2	1.07	1.1	1.06
Leadership position					3.3**	1.35	3.3**	1.35	3.2**	1.36
Party in government					-3.0*	1.47	-2.9*	1.52	-2.9*	1.73
Share of seats held by party					-1.1	3.90	-0.42	3.99	-2.8	4.35
Party ideology left to right (scaled)					0.92	0.618	1.1	0.706	1.8	1.11
N	2,094		2,094		2,094		2,094		2,094	
Adjusted R ²	0.016		0.016		0.021		0.021		0.021	

¹p<0.1; ²p<0.05; ³p<0.01; ⁴p<0.001

²SE = Standard Error

A.4.3. Additional populism analysis

Populism correlations (party level)

	Party ideology left to right	Populism	Intra-party democracy	Nativism	Complex versus common-sense politics
Party ideology left to right		0.23	-0.66	0.73	0.23
Populism	0.23		-0.52	0.66	0.88
Intra-party democracy	-0.66	-0.52		-0.74	-0.49
Nativism	0.73	0.66	-0.74		0.54
Complex versus common-sense politics	0.23	0.88	-0.49	0.54	

Intra-party only regression models

DV:# of missing tweets (log)						
Variable	M1		M2		M3	
	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²
Intra-party democracy (scaled)	-0.02	0.023	-0.02	0.023	-0.02	0.029
N tweets in 2018 (log)	0.91****	0.015	0.91****	0.015	0.91****	0.015
Male			0.03	0.032	0.03	0.033
Minister (ever)			-0.06	0.043	-0.07	0.044
Leadership position			-0.07	0.088	-0.06	0.090
Terms in parliament			0.02*	0.012	0.03**	0.012
Party in government					0.08	0.062
Share of seats held by party					0.02	0.190
Party ideology left to right (scaled)					-0.01	0.034
Adjusted R ²	0.869		0.869		0.869	

¹*p<0.1; **p<0.05; ***p<0.01; ****p<0.001

²SE = Standard Error

Nativism only regression models

DV: # of missing tweets (log)						
Variable	M1		M2		M3	
	Beta ¹	SE ²	Beta ¹	SE ²	Beta ¹	SE ²
Nativism (scaled)	0.05*	0.025	0.04	0.027	0.07*	0.043
N tweets in 2018 (log)	0.91****	0.016	0.92****	0.016	0.92****	0.016
Male			0.02	0.033	0.02	0.033
Minister (ever)			-0.06	0.044	-0.07	0.044
Leadership position			-0.07	0.089	-0.06	0.091
Terms in parliament			0.02*	0.013	0.02*	0.013
Party in government					0.06	0.057
Share of seats held by party					0.11	0.199
Party ideology left to right (scaled)					-0.05	0.039
Adjusted R ²	0.869		0.869		0.869	

¹*p<0.1; **p<0.05; ***p<0.01; ****p<0.001

²SE = Standard Error