

# Understanding Privacy Switching Behaviour on Twitter

Dilara Keküllüoğlu  
University of Edinburgh  
Edinburgh, UK  
d.kekulluoglu@ed.ac.uk

Kami Vaniea  
University of Edinburgh  
Edinburgh, UK  
kvaniea@inf.ed.ac.uk

Walid Magdy  
University of Edinburgh  
Edinburgh, UK  
wmagdy@inf.ed.ac.uk

## ABSTRACT

Changing a Twitter account's privacy setting between public and protected changes the visibility of past tweets. By inspecting the privacy setting of more than 100K Twitter users over 3 months, we noticed that over 40% of those users changed their privacy setting at least once with around 16% changing it over 5 times. This observation motivated us to explore the reasons why people switch their privacy settings. We studied these switching phenomena quantitatively by comparing the tweeting behaviour of users when public vs protected, and qualitatively using two follow-up surveys (n=100, n=324) to understand potential reasoning behind the observed behaviours. Our quantitative analysis shows that users who switch privacy settings mention others and share hashtags more when their setting is public. Our surveys highlighted that users turn protected to share personal content and regulate boundaries while they turn public to interact with others in ways the protected setting prevents.

## CCS CONCEPTS

• **Security and privacy** → **Social aspects of security and privacy**; • **Human-centered computing** → *Social content sharing*; *Social media*.

## KEYWORDS

Privacy, Security, Online Social Networks, Twitter, Privacy Settings

## ACM Reference Format:

Dilara Keküllüoğlu, Kami Vaniea, and Walid Magdy. 2022. Understanding Privacy Switching Behaviour on Twitter. In *CHI Conference on Human Factors in Computing Systems (CHI '22)*, April 29-May 5, 2022, New Orleans, LA, USA. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3491102.3517675>

## 1 INTRODUCTION

Interpersonal boundaries are continuously shifting concepts for people as they move between situations, contexts, and as the world they exist in changes. Humans are social creatures and are well accustomed to this continuous shifting of how they present themselves and how they manage their boundaries around personal privacy, self presentation, and access to self. The introduction of social

media platforms though can lead to challenges in boundary management where many of the natural high-granularity actions taken in the physical world must be translated into the more rigid privacy setting options used by platforms [12]. While many people either do not change the default settings provided by the platforms [13] or set them only once during account creation [47], other people have found a myriad of ways to use social media platforms' rigid setting structures to mimic the fine gradations of inter-personal boundaries, often in unexpected and non-obvious ways. In this work, we look at Twitter's tweet visibility setting feature which is particularly ridged with only two options, public and protected, which apply to the user's entire tweet stream, including historical tweets. Yet even in this highly binary and low fidelity situation we see a range of how people use the system to create a rich boundary management to suit their needs.

Twitter is a particularly interesting platform to study in regards to privacy management because of the limited control options and also because Twitter is quite blunt in its help center about the impact of switching from protected to public: "Unprotecting your Tweets will cause any previously protected Tweets to be made public" [7]. The implication is that Twitter accounts are meant to be public or protected and are generally assumed to stay in one of those states with changes between being rare. Existing work on privacy settings on Twitter mostly focus on the differences between users who are protected and public [8, 29]. These studies check the settings of user accounts only once to decide if it is public or protected. However this assumption may be wrong and it may be the case that users are changing their account visibility often enough to count as having a more complicated visibility status.

Studies in other platforms, like Facebook, have found that even when privacy controls do not well match the effects users are trying to create, people are quite good at using the technology in unexpected ways to create the effects they want. A good example is work on how teenagers manage context in social networks by doing things like deactivating their accounts when they are not logged in [36]. While not immediately intuitive, deactivating an account prevents others from interacting with it, enabling the user to only allow others to interact with the account when they themselves are logged in. The point is that privacy permission management is not necessarily a single set-it-and-forget-it setting. People use these settings in boundary management, which is a continuously changing state between people.

While users may need to manage interpersonal relationships with high granularity, the main control over privacy on online social networks is still the visibility of the account content. On Twitter, that means the visibility of the account and its tweets. Twitter users have a relatively small set of options to protect their tweets and protected user accounts also face restrictions on the types of interactions they can have. For example, if a protected user

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9157-3/22/04...\$15.00

<https://doi.org/10.1145/3491102.3517675>

mentions a non-follower, that person cannot see the mention and therefore cannot respond to it. Other than changing tweet visibility, users can also delete their tweets, block other users to prevent them from interacting, and deactivate their accounts.

It is also an open question what effect users are attempting to create when they change visibility settings. The most obvious assumption is that they are trying to hide their tweets from public consumption. But Twitter also has other issues, such as harassment [22], that may cause users to switch to protected [51]. This observation opens an interesting question about how users behave when they are public vs when they are protected. A user simply trying to avoid harassment may actually not change behaviour when they are protected, because it is themselves they are trying to protect, not the tweets. Users that are trying to protect their tweets may also have specific types of tweets that they are more inclined to post while protected as opposed to when public.

In this work, we focus on users who change their privacy settings in order to understand the behaviour and motivations behind the changes, as well as the differences in self-disclosure between when they are public versus protected. More precisely, we investigate the following research questions:

- RQ1** How frequently do Twitter users change their tweet visibility settings?
- RQ2** Do users employ different posting strategies when they are public vs protected?
- RQ3** Why do users change their visibility settings and how do the reasons differ between changing to protected versus public?
- RQ4** What other strategies do switching users employ to control their audience and interactions?

To answer these questions, we curated a set of 107K users, whose accounts were protected at the start of the data collection, and analyzed their account visibility changes for three months. Of these, 45K changed their visibility to public at least once allowing us to collect their tweets to determine if their tweeting behaviour changes when they were protected compared to when they were public. We also conducted two user surveys to get insights into why users decide to change their tweet visibility back and forth.

Our findings show that a large portion of protected users do change their privacy setting to public, around 40% of the protected accounts we inspected changed to public at least once within three months. A quarter of those accounts changed more than ten times within this period. Analysis of tweet data shows that users' accounts have less tweets during times when they are protected. They also mention other users less compared to when they are public. Our survey results show that users mostly change their tweets to protect to regulate their boundaries but change to public to overcome the platform's technical constraints. Our findings provide a unique perspective into the usage of privacy settings in a platform where options are binary and in an account-level, where trade-off between privacy and functionality must be calculated with the historical posts in mind. We also provide design implications built upon our findings and prior literature to help users and platforms minimize potential privacy leaks.

## 2 RELATED WORK

People use online social network platforms in a range of ways including connecting with others, gaining social capital, and learning about events, all while also wanting to protect their privacy [12, 53]. People also use social media to seek support [4, 11], build reputation [49], and for professional contexts [34, 35]. People can choose to only connect with a specific group of people, or multiple social circles. This usage can cause context collapse where these social circles, such as friends, family, and colleagues, are all on the same social network [35, 41, 52] making it hard for users to control the flow of information.

Altman [3] describes privacy as a dynamic process where people select what part of themselves they want to show, when they want to show it, and who they want to show it. It is not trivial to regulate this on social media where posts are usually permanent [33]. Especially on Twitter where privacy settings are not granular and once changed they apply to all of the posts shared in the past. Once shared, the information becomes co-owned with the selected audience [42]. Hence, it is critical to select what information to impart and the audience that will see it on social media.

### 2.1 Usage of Privacy Settings

Social network platforms provide privacy settings for users to adjust a range of configurations including setting the audience of their posts, regulating interactions, changing the visibility of shared information, and so on. However, users keep the defaults provided by the system [13, 18, 31] for various reasons such as difficulty of finding and using the settings [45, 50], uncertainty about their preferences [1], or not trusting the system to keep the configured settings [45]. Some users choose not to change their settings even when there are posts shown to be violating their privacy [32, 38] and some configure their settings only during initial setup [47].

Conversely, some users utilize privacy configurations given by platforms to protect their privacy online [12, 33, 36]. However, configuring privacy settings to accurately reflect the user's intended audience is not trivial [20, 30, 33]. Privacy violations in social media may be caused by unclear permissions or people not knowing how the permissions work. Liu et al. [30] analyzed the privacy settings set by Facebook users and compared them to the desired settings to understand the discrepancies. They found that users' expectations did not reflect the actual settings of their posts most of the time and that the size of their audience was also larger than expected.

Fiesler et al. [13] compares public and private Facebook posts to analyze what information is shared and who they are shared with. They collect recent posts of 1,815 Facebook users and analyze the contents as well as their privacy settings. They find that public posts and non-public posts have similar content but users with certain demographics tend to utilize the settings. Liang et al. [29] compares public and protected users in Twitter and find that users who use privacy settings are more likely to disclose geo-location in their tweets.

Privacy settings provided by platforms have different granularities. For example, Facebook allows users to configure each post's audience where in Twitter the settings applied to the whole account. Hence, users might utilize them in different ways. Choi and Bazarova [8] compares users' behaviour on Facebook and Twitter.

They find that while users share more intimate content on Facebook, there are no differences between protected and public Twitter users in terms of content. Similarly Liang et al. [29], find that public Twitter users had more privacy concerns than the protected ones since the latter have more clear and less permeable boundaries.

Most prior work compares public and protected user behavior using a between-subjects design. That is, they look at what users who are private and users who are public do. These studies typically find that public and protected users generally share similar things [8, 13]. However there is an obvious gap here as it may be that the same user switching between public and protected does not share in a similar way. In this work, we use a within-subject design where we consider how behavior changes when a user changes their visibility.

## 2.2 Controlling Content and Interactions

Other than configuring the privacy settings, users utilize various methods to protect their privacy online including self-censorship [9, 35, 45], limiting connections [12, 23], creating multiple accounts for different purposes [48, 53, 56], as well as deactivating [36]. Users sometimes take drastic measures such as stopping usage completely [5, 17, 27] to protect their privacy, which can have effects on their social connections [17], and professional life [6]. Cutting off connection can also be harmful for vulnerable populations [11]. Lampinen et al. [28] interviewed 27 participants about online sharing behaviours and boundary regulation tactics. They divide their identified set of strategies into *preventative* and *corrective*. Preventive strategies included not sharing particular contents, or sharing content only with a targeted audience. Corrective strategies included deleting an already shared post or presenting it as a joke.

Online content and interactions has a persistence over time that is not present in offline interactions. Users can choose to delete some or all of their posts for various reasons such as regret [46], typos [2], to prevent temporal context collapse [21], and so on. However, the residual activity around the deleted/hidden tweets can still be used to infer the content [2, 25, 26, 37]. Almuhimedi et al. [2] followed 292K users for a week to analyze deleted tweets. They found that around half of the users deleted at least some tweets during the week. The content of the tweets that were deleted were not substantially different than the ones that were left on the platform. Mondal et al. [37] analyzed the accessibility of tweets longitudinally by trying to recollect tweets from older datasets that date back to six years. Unlike Almuhimedi et al. [2], they also included the tweets of users who turned their accounts to protected to their analysis. They found that while most of the users did not withdraw any of their tweets, 8.3% of them deleted some tweets, 16% deleted their account, and 10% protected their tweets. They found that they can recover the topics of interests of the withdrawn tweets by analyzing residual interactions around them which shows that even deletion might not prevent possible privacy risks.

Users also try to sanction interactions without harming their relationships to regulate boundaries. Rashidi et al. [44] interviewed 23 young adults to analyze the sanctions enforced by them on popular social media platforms. They grouped the sanctions using three dimensions: on-site and off-site, individual and collaborative, visible and invisible sanctions. They found that people prefer using

invisible sanctions like muting rather than blocking, unfriending, or deleting content.

Instead of sanctioning, users may choose to deactivate their accounts to protect their privacy, concentrate on their work [5], and to limit interactions [36]. Platforms provide deactivation as an alternative to deleting accounts and users are allowed to get their accounts back if they decide to do so. For example, Twitter gives a 30 day grace period during which users can reactivate their deactivated accounts. After that time the accounts will be permanently deleted and the username can be claimed by someone else. Ng [40] describes the users who temporarily deactivated their accounts as intermittent discontinuers and found that social media fatigue was one of the reasons of the deactivation. They also found that some users change their tweets to protected instead of deactivating.

## 3 MONITORING SWITCHING BEHAVIOUR ON TWITTER

To understand the phenomena of users switching their privacy settings on Twitter, we initially curated a large set of users who had protected accounts at the start of the data collection and monitored their behaviour over three months to record any possible switch to their status between protected and public. We analyzed these switching patterns and compared their tweets using features such as sharing of hashtags, media, links, and so on. In this section, we describe the Twitter API as well as our data collection and analysis.

### 3.1 Ethical Considerations

We understand that even though we access the tweets of public accounts, users may have more fluid account types as shown in this study. While Twitter is quite clear about the implications of making an account public and most of the survey participants evidenced awareness that their tweets can be accessed by non-followers when they changed their tweets to public, it might not be clear to the users that researchers are also included on that set [14]. To mitigate possible risks, we only report on the metadata curated from the social media data. We do not single-out any users and avoid sharing their social media content to protect their privacy. We also do not collect social media usernames or data of the survey takers. Only the researchers on this project have access to the collected raw data. We followed our University's ethics protocol in the design and running of this work, including the social media data collection and following surveys.

### 3.2 Twitter Functionality

Twitter has a binary privacy configuration where accounts are either public (default) or protected. All tweets associated with a *public* account can be seen by anyone on the Internet, while all tweets associated with a *protected* account can only be seen by the followers of that account. Changing an account's setting changes the visibility of all tweets associated with this account, including past tweets. Any Twitter user can follow a *public* account, while following a *protected* one first requires the account owner's approval. A *protected* account's tweets can be liked by the followers but they cannot be retweeted or quote tweeted. Interactions of *protected* accounts can only be seen by the followers. Hence, if these *protected* accounts reply to a public tweet of a non-follower, that will not be visible to

that non-follower. This also applies to mentioning users by using their handles (@username), quote tweeting them, as well as liking and retweeting them. Hence, *protected* accounts must change to *public* if they want their interactions with non-followers to be seen, but doing so also makes their full tweet history public. On the other hand, if a *public* account mentions a *protected* one, anyone on the Internet can see the content since the owner of the tweet is *public*. We use tweet visibility and account visibility interchangeably in this paper.

### 3.3 Collecting Switching Users

As mentioned, Twitter accounts are by default public and users need to actively change their account to protected if they want to limit their tweet visibility. Since our aim is to study users who switch between public and protected, we collected users with protected accounts which indicate that they have changed their tweet visibility at least once. To obtain our sample of protected users, we first sampled the Twitter stream without any filters and collected a list of mentioned protected users from these tweets for 60 hours starting on the 29th June 2020. During this period, we managed to collect tweets that contain mentions to 3.15M users. We inspected the account visibility of each of those users using the Twitter user API, and found that around 4% of them were protected accounts at the time of the data collection.

Our initial pool of protected users contain around 107K accounts. We applied a frequent check to get the account visibility of these protected accounts every 30 minutes for three months from 17th July until 17th October 2020<sup>1</sup>. Out of the 107K users that were protected on the first check, 64.5K (~60%) of them stayed protected during all three months. The remaining 42.5K (~40%) have switched their account visibility to public at least once during this period, including 7.6K (7.1%) who changed to public and stayed that way. Table 1 shows the number of switches users made in the three months we collected their account visibility.

A user changing their tweet visibility settings frequently does not necessarily mean they stay protected/public equally. They can predominately stay protected or public and only switch for short time periods. Hence, we calculated how long our users stayed protected/public during the three months. For each user, we compared the number of 30 minutes blocks when they were protected to the number of blocks they were public. 14.6K (34.5%) of the users who changed their account visibility at least once preferred to stay protected more than 95% of the time. More than half of the switchers stayed protected 80%+ of the time.

### 3.4 Collection and Labeling of User Tweets

We collected the tweets of the switching users when their account was set to public. While it is possible to collect the last 3200 tweets of a user via Twitter API, we only collected the tweets posted during the account type check period (i.e. between 17th July and 17th October 2020). We managed to collect the tweets of 25.6K (60.4%) users out of the 42.5K switching accounts in our collection. Some of these users change their account visibility frequently and stay

<sup>1</sup>During this period, there were some instances where we could not reach some users because they were deactivated, suspended, or there were temporary problems with the Twitter API. There was only one instance where we could not reach any accounts because of a service disruption.

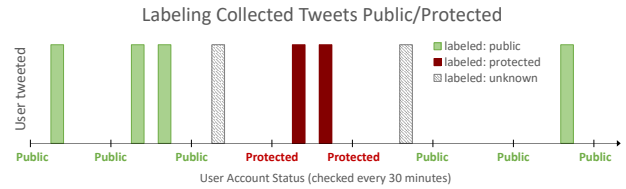


Figure 1: Example of the labeling process of the tweets of users in the dataset.

public for a short time. Hence, we could not get any tweets from the remaining 16.9K users.

After collecting the tweets, and since we have the visibility status of the account every 30 minutes, we labeled them according to the collected account visibility of the user when the tweet was posted. If a tweet was sent at time  $t$ , we compared the user’s account visibility collected just before and after  $t$ . If the account visibility settings were the same, we gave the tweet that label. In cases where the account setting visibility was different or we could not retrieve one of them, we discarded those tweets and did not include them in our analysis. Figure 1 illustrates an example of the labeling process of the tweets in a users’ timeline. For simplification purposes, we call tweets we labeled as public  $t_o$  and the ones we labeled as protected  $t_x$ . Table 2 gives the notations and definitions for the tweets we labeled. Ultimately, we were able to label nearly 39.2M tweets (including retweets and replies) collected from the 25.6K switching accounts in the three months, where 26M (66.2%) of them are  $t_o$  (public) while the remaining 13.2M (33.7%) are  $t_x$  (protected). Nearly 9M (23%) of these tweets were labeled as English by Twitter, and the remaining 77% covered 64 other languages including Japanese, Portuguese, Indonesian, and Korean in that order.

Table 3 shows the user counts by percentages of the time the account was protected during the three months, as well as the number of total tweets we collected from these users. There is a positive relation between the time users stay protected and the number of tweets they send while they are protected. However, some users tend to tweet less when they were protected even if they stayed protected most of the time.

### 3.5 Comparing tweets characteristics of users

To understand the reasons why users switch their account visibility, we divided them into three groups based on their privacy settings usage and analyzed them separately. These groups are  $u_x$ , the users who stay *protected* 90% of the time or more,  $u_o$ , the users who stay *public* 90% of the time or more, and  $u_b$ , the remaining users who have more *balanced* duration of staying public or protected compared to  $u_x$  and  $u_o$ . In this section, we compare 11 features of the  $t_x$  and  $t_o$  tweets of  $u_x$ ,  $u_b$ , and  $u_o$ . Table 2 shows the above notations and definitions for easy reference.

As shown in Table 3, out of the users we collected tweets from, 3K (11.7%) of them were  $u_x$ , 20.1K (78.3%) of them were  $u_b$ , and 2.6K (10%) of them were  $u_o$ . Users in our dataset have the majority of their tweets with the same visibility state as the dominant setting.

	0	1	2	3 - 5	6 - 9	10+	Total
# Users	64,460	7,630	9,195	8,599	5,910	11,121	106,915

**Table 1: Number of switches users made in three months.**

Notation	Definition
$t_o$	Tweets labeled as <b>public</b> by our system, i.e. user was public in our account type checks preceding and succeeding the tweet.
$t_x$	Tweets labeled as <b>protected</b> by our system, i.e. user was protected in our account type checks preceding and succeeding the tweet.
$u_o$	Users who stayed public 90%+ of the time we collected the account types.
$u_b$	Users who stayed protected between 10% and 90% of the time we collected the account types.
$u_x$	Users who stayed protected 90%+ of the time we collected the account types.

**Table 2: The notations and the definitions we use for users and labeled tweets of the Twitter data collection.**

% time as protected	[0-10]%	(10-50)%	(50-90)%	90+%
# Users	2.6K	9.8K	10.3K	3K
# $t_o$	5.3M	14.3M	6M	375K
# $t_x$	214K	3.4M	7.1M	2.5M
# $t_o + t_x$	5.5M	17.7M	13.1M	2.9M

**Table 3: User counts by the percentage of time the account was protected during the three months of data collection. Also the total number of  $t_o$  and  $t_x$  collected for each group of users. Only the users we could collect tweets from are reported.**

Here we compare the tweeting frequencies, language-independent tweet features, and English tweeting behaviour of the three user groups when they tweet under public or protected settings. Looking at tweet features can give us insights on why users might change their visibility settings back and forth. Hence, for each tweet we look at the following features: mentions (tweet has at least one mentioned username), verified mentions (mention of a verified user), non-follower mentions, reply (the tweet is a reply to another existing tweet), retweet (RT), quote tweet (QT), URLs, hashtags, media (e.g. photo and video), and whether the tweet is in English using the language label given by the Twitter API. We did not take the geo-location as a feature since only a small portion of the users had it enabled. We collected the user profiles of the mentions to check whether they were verified. We also collected the followers of our users to check whether they mention non-followers in their tweets.<sup>2</sup> We extract these features from the tweets of  $u_x$ ,  $u_b$ , and  $u_o$ . We then calculate the percentage of prevalence in  $t_x$  and  $t_o$  for each of the features and compare them for every user. We also divide the number of tweets a user sent while public/protected by the duration of staying public/protected so we can calculate the tweeting frequencies. We compare these features using paired sample t-test and apply Bonferroni correction post-hoc with p-values divided by 14.<sup>3</sup> For this analysis, 168, 675, and 227 users from  $u_x$ ,  $u_b$ , and  $u_o$  respectively were excluded due to all of their tweets being either public or protected.

<sup>2</sup>The data collection for both of these features was done at a later time than the initial data collection. Considering Twitter networks are dynamic, the state of the network might have changed between these two collection times.

<sup>3</sup>We made at most 14 comparisons for each user group but only report 11 of them in this paper.

Table 4 shows the means of values, 95% confidence intervals, and paired t-test p-values for each feature of the tweets for  $u_x$ ,  $u_b$ , and  $u_o$ . Users in  $u_x$  tweet more when they are public. They also mention Twitter users in their tweets more when they are public, albeit with a low effect size. The effect sizes are bigger for mentioning verified users and non-followers. Since a user's tweets can only be seen by followers, it is not surprising that they are switching to public to mention a verified account or someone who does not follow them. Users in  $u_x$  also share more tweets with hashtags in them when they are public. They share media and links along with their tweets, reply, retweet, and quote tweet in similar rates. The effect sizes for comparisons of tweeting frequency, verified user mentions, and non-followers mentions were small and the remaining comparisons all had very small effect sizes.

Similar to  $u_x$ , users in  $u_b$  tweet and mention other Twitter users in their tweets more while in the public setting. They also share media, links, and hashtags in their tweets and reply more when public. On the other hand, they retweet more when they are protected. Surprisingly, they mention verified users and non-followers in similar rates. They also tweet in English and quote tweet similarly when public and protected. Aside from the tweeting frequency, which had a small effect size, effect sizes were all very small for comparisons of  $u_b$ .

Users in  $u_o$  also tweet more when they are public. They mention users, reply, and share hashtags more when they are public. However, surprisingly, they mention verified users and non-followers relatively more when they are protected. They tweet in English, retweet, quote tweet, share media, and links in similar rates. The effect sizes for comparisons of  $u_o$  were all very small.

		$\bar{t}_x$	95% CI	$\bar{t}_o$	95% CI	p-value
$u_x$	Tweet Frequency	0.22	[0.21, 0.23]	0.59	[0.54, 0.63]	***<.001
	Mentions	61.79	[60.99, 62.59]	64.11	[63.19, 65.02]	***<.001
	Verified Ment.	10.34	[9.63, 11.05]	15.30	[14.25, 16.35]	***<.001
	Non-follower Ment.	52.28	[50.34, 54.23]	56.80	[54.66, 58.94]	***<.001
	Reply	42.43	[41.48, 43.39]	43.29	[42.23, 44.35]	.022
	RT	22.59	[21.68, 23.51]	22.02	[21.07, 22.96]	.058
	QT	8.15	[7.80, 8.50]	8.39	[7.98, 8.81]	.169
	Urls	7.99	[7.68, 8.30]	8.49	[8.07, 8.92]	.013
	Hashtags	3.70	[3.43, 3.97]	5.60	[5.15, 6.05]	***<.001
	Media	15.12	[14.61, 15.63]	15.07	[14.48, 15.65]	.835
English	17.06	[15.98, 18.14]	17.28	[16.17, 18.39]	.178	
$u_b$	Tweet Frequency	0.28	[0.27, 0.28]	0.50	[0.48, 0.51]	***<.001
	Mentions	64.08	[63.77, 64.38]	65.73	[65.45, 66.02]	***<.001
	Verified Ment.	11.00	[10.76, 11.25]	11.10	[10.86, 11.34]	.228
	Non-follower Ment.	57.16	[56.62, 57.71]	57.41	[56.89, 57.93]	.097
	Reply	41.31	[40.94, 41.68]	41.73	[41.38, 42.08]	***<.001
	RT	25.57	[25.20, 25.93]	25.19	[24.85, 25.53]	***<.001
	QT	8.53	[8.40, 8.66]	8.61	[8.48, 8.73]	.065
	Urls	8.36	[8.24, 8.49]	8.69	[8.56, 8.82]	***<.001
	Hashtags	4.19	[4.07, 4.31]	5.16	[5.03, 5.29]	***<.001
	Media	15.47	[15.27, 15.67]	15.80	[15.61, 15.98]	***<.001
English	19.55	[19.11, 19.99]	19.63	[19.19, 20.07]	.090	
$u_o$	Tweet Frequency	0.47	[0.44, 0.50]	0.54	[0.51, 0.57]	***<.001
	Mentions	66.43	[65.41, 67.44]	68.64	[67.86, 69.41]	***<.001
	Verified Ment.	15.78	[14.69, 16.86]	13.36	[12.50, 14.23]	***<.001
	Non-follower Ment.	57.35	[55.89, 58.80]	55.20	[53.96, 56.45]	***<.001
	Reply	40.35	[39.19, 41.52]	42.45	[41.47, 43.44]	***<.001
	RT	27.35	[26.20, 28.50]	26.78	[25.80, 27.77]	.093
	QT	8.82	[8.33, 9.31]	8.94	[8.60, 9.28]	.522
	Urls	8.82	[8.32, 9.32]	8.91	[8.57, 9.25]	.673
	Hashtags	4.82	[4.36, 5.28]	5.50	[5.14, 5.87]	*.001
	Media	16.99	[16.25, 17.72]	16.13	[15.62, 16.64]	.004
English	23.05	[21.62, 24.48]	23.36	[21.98, 24.74]	.127	

**Table 4: Paired samples t-test comparing number of tweets of  $u_x$ ,  $u_b$ , and  $u_o$  when they are protected and public (with Bonferroni correction \* $p < (0.05/14)$ , \*\* $p < (0.01/14)$ , \*\*\* $p < (0.001/14)$ ).**

## 4 USER SURVEYS

The above data collection and analysis provides an interesting view on *what* users did when they switch account visibility, but it does not provide an understanding of *why* they are engaging in these actions. In this section we describe two surveys that we ran to better understand the reasons users change their settings and how common those reasons are. We start with an initial open-ended survey where we elicit reasons why people have switched in the past. We then combine the results of that survey with related work and our own observations during the data collection detailed in the previous section to create a list of common reasons people might change their visibility settings. We use the list in a second closed-ended survey to understand how common different switching reasons are. We give the full list and the sources of each reason in Appendix A.

### 4.1 Identifying reasons to switch account visibility

The first survey aims to identify common reasons that lead users to switch between public and protected. The survey is not intended to identify a comprehensive set of reasons for switching, and instead

focuses on identifying common reasons which are then combined with information from literature (e.g. [36, 44, 51]) to construct the options used in the next survey.

We recruited 100 Twitter users who have switched their tweet visibility at least once in the prior two months. The study was advertised as a “Short survey about changing your tweets between public and protected on Twitter” to Prolific Academic (PA) [43] users who are fluent English speakers on March, 2021. We conducted two pilots surveys with six participants each to adjust the wording of the questions as well as to get an accurate estimation of the time required. 100 participants took the final survey with an average completion time of 5.5 minutes and a compensation of £1.15. The study design was approved through our University’s ethics process.

*Survey Instrument:* The survey consisted of: informed consent, a screening question about if they had changed their Twitter account visibility (the ad clearly stated that this was required), four retrospective questions described below, demographics questions which included Twitter usage questions, and an optional free text comments box. For the retrospective questions we asked them to “think back to a recent time when you have changed your Tweets from public to protected or from protected to public.” The first question

Theme	# users	Theme	# users
Sensitive/Political/Controversial	19	Giveaways*	4
To prevent people I know from seeing	16	Harassment	3
Personal Information	15	Avoid suspension	3
To prevent non-followers from seeing	10	Viral Tweet	3
Sense of privacy	10	Archival purposes	2
Broader audience*	7	NSFW	1
Interacting with non-followers*	4		

**Table 5: Themes of free-text answers about reasons of switching. Most reasons are for switching from public to protected, while (\*) indicates reason to switch from protected to public.**

was free text and asked them what had motivated them to make the switch, followed by another free text question asking them what they hoped to achieve by making the switch. We also asked if the switch enabled them “to achieve the effect I was trying for” (Likert) and what direction the switch was in (to public, to protected, or multiple changes).

*Analysis:* Two researchers reviewed the responses and determined that the two free-text answers were often very related to each other, and that the answer for what they hoped to achieve often was an expansion of the motivation for the switch, such that viewing them together provided a more comprehensive understanding of the reason. The answers were therefore analyzed together. We used an affinity diagram [19] type approach to conduct a thematic analysis. The answers were placed in a shared spreadsheet, one researcher went through and sorted the answers into themes. They then met and the second researcher read through all the answers and adjusted the themes discussing with the first researcher as they went. The result was a set of themes that both researchers agreed on (Table 5).

*Participants:* Participants were 23.6 years old on average with a max age of 39 years. 52 identified as male and 48 female. 65 described a situation where they switched to protected, 11 described switching to public, and 24 described an event that required multiple visibility changes. When asked about the normal visibility of their account, 42% indicated that they kept their account public most of the time, 27% kept it mostly protected, while 31% indicated that they switched more often. 62% had also tweeted within the last week.

**4.1.1 Reasons to Switch:** Participants most commonly switched to protected to talk about sensitive, political, or controversial topics freely. Reasons in this theme also overlapped with other themes, particularly around avoiding problematic interactions with strangers and proactively limiting the audience. Another common reason was to prevent the people they know, such as friends and family, from seeing their tweets. Some of these users were uncomfortable because a person they knew found their account leading them to switch visibility. Other participants were uncomfortable with strangers or non-followers seeing their tweets. Participants also changed to protected temporarily to share personal information such as private Instagram links and then deleted the tweet before switching back to public. Harassment, having a tweet go viral, and sharing tweets that were not safe for work (NSFW) were some of the other reasons given. Uncommon but interesting reasons included wanting to archive an account and to avoid getting suspended due to complaints.

Our participants primarily changed to public to reach a broader audience, interact with non-followers, as well as participating in giveaways where the account must be public to get selected for the prize. One participant cited all of these reasons when explaining why they changed to public. “*I hoped to achieve what the platform could offer to its fullest, that included retweeting thing in order to comment on them, “taking part” in a trending hashtag or even taking part in giveaways.*” Interestingly, one participant discussed going public for a while in order to gain more followers, then switching back to protected. Others said that it was challenging for friends and family to find them when they were protected, so they went public for a while to make themselves easier to find. We also asked if changing visibility enabled them to achieve the effect they were trying for and 89% of the participants agreed that it did.

## 4.2 Prevalence of Reasons

The second survey looked at how common different reasons for switching visibility are. We used a combination of the results of the first survey, our observations from the Twitter data collection, and findings from related work to construct lists of reasons people switch to public and to protected. Then used the survey to find how many participants have switched due to these reasons.

We advertised a study entitled “5 min survey about changing your tweets between public and protected on Twitter” on PA to fluent English speakers on April, 2021. The advertisement stated that we were looking for people who had changed their tweet visibility between public and protected two or more times in the last year to ensure respondents had experience with switching. We conducted a pilot survey with five participants to ensure accurate time estimates. 324 participants took the survey, requiring an average of 4 minutes to complete. They were compensated £0.75. Ethics approval was given by our University.

*Survey Instrument:* Participants were first shown a consent page, followed by a screening to see if they had switched two or more times in the last year, followed by three multiple answer questions described next, questions about their Twitter experience, demographics, and an optional comment box. The first two multiple answer questions asked users to indicate “which of the following has previously lead you to change your tweet visibility settings to public?” and also for protected. The third multiple answer question showed a list of protection activities, like deleting tweets when moving from protect to public, and asked users to select those they had used previously. We followed with a scenario-based question

about a person who tweeted while protected and then switched to public and asked if their tweet was now public or protected.

*Participants:* Participants were an average of 25 years old ( $\sigma = 7.6$ ), with 46% identifying as female, 53.4% as male, and the remainder preferring not to say.

Participants were asked what their normal public/protected balance was on their account. Ninety-nine (30.5%) indicated their account was “Always” or “Mostly” public (Public here forth), 28.4% indicated their account was “Always” or “Mostly” protected (Protected), and finally 41.0% indicated that their visibility was “Balanced” or “Somewhat” public/protected who are referred as Balanced users for the remainder of this section.

The majority (67.9%) of participants accurately answered our scenario-based question about tweet visibility and said that the tweet sent while protected would now be publicly visible since the person switched to public. Fifty-three (16.4%) thought that only users logged into Twitter could see it, but were aware that the tweet would be public to all Twitter users. But 15.7% incorrectly thought that the tweet would remain protected.

We asked participants how often they had changed their visibility in the last three months. Sixty-two (19%) had not changed their tweet visibility settings in the last three months. One hundred and sixteen (35.8%) had changed once, 27.5% twice, 15.4% three to five times, and 2.2% changed six or more times.

**4.2.1 Results:** We asked participants to select all of the reasons that have caused them to change their visibility settings to public (Table 6) and protected (Table 7). On average users selected 3.3 ( $\sigma = 2.1$ ) protected reason options and 3.4 ( $\sigma = 2.0$ ) public reason options. The small number of selections and the range of selections shown in the table suggest that respondents were indeed reading the choices and only selecting those appropriate to them.

The most common reason to turn public for those that were mostly public or balanced was to “reach a broader audience and get more interactions”. The finding makes sense given that interacting with tweets is a key functionality of Twitter. The most common set of reasons to turn public was to be able to interact with other accounts through mentioning, replying, retweeting, and quote tweeting. For those who were mostly protected, interactions like these seem to be the main driver to change visibility.

The most common reason to turn protected for all groups was “to prevent non-followers from seeing tweets with personal content”. In the first survey we saw several people switching to protected because someone they knew had found their account, and in this survey we similarly see that being a large reason to switch, even more so than preventing friends and family from seeing tweets, suggesting that the issue is around specific individuals more than just people they know. Preventing interaction from strangers and harassment were also common answers. “To get a sense of privacy” was also a common answer, though interestingly not the most common. In the first survey, a common vague answer was that the user just wanted more privacy, which is why we added the answer option to the second survey despite its low specificity. Interestingly, sharing content like pictures, links, and not safe for work content were not common reasons to turn protected. Similarly, interacting with other users, which was a common reason to turn public, was not a common reason to turn protected.

We also examined the combinations of answers users gave across these questions around turning public and protected. The most frequent pair was “To reach a broader audience and get more interaction with my tweets” and “I wanted to prevent non-followers from seeing tweets with personal content” with 30.2% of participants giving both those answers. Similarly, 18.5% both indicated turning public to get more interactions while also indicating that they turned protected to prevent interactions from strangers. Another pair selected by 11.4% participants was “To enter to get giveaways or freebies” and “I wanted to prevent non-followers from seeing tweets with personal content”.

We also asked participants to select actions they have taken to control who can see and interact with their tweets (Table 8). More than half indicated that they delete some or all of their tweets before changing their visibility to public. The control approaches of soft blocking, muting, and blocking were also used by many. Majority protected participants in particular preferred soft blocking, possibly because it removed followers without creating a notification to the impacted Twitter account and may be seen as less harsh.

Interestingly, 23.1% of participants indicated either temporarily deactivating their accounts to prevent interactions or changing to protected when they are not logged in and cannot respond to interactions. Twitter allows 30-day grace period for accounts that are deactivated, if a user re-activates their account in this time frame all content will be restored. We added these options based on prior research showing that some Facebook users delete their accounts when not online as a way of preventing interactions when they are unable to respond [36]. Based on these results, some Twitter users may be using a similar tactic. Temporarily deactivating might also explain why we could not reach some accounts during the automated Twitter data collection.

## 5 DISCUSSION

### 5.1 Summary of Findings

We quantified privacy settings usage on Twitter to answer our RQ1 “How frequently do Twitter users change their tweet visibility settings?” After monitoring a set of 107K protected accounts on Twitter over three months, we found that 40% of these accounts changed their visibility to public at least once over this period. Over a quarter of those who changed, did so 10 or more times. This shows an interesting behaviour of Twitter users when it comes to their use of tweet visibility settings, which motivated our RQ2 related to the posting strategies of these users when they are public vs protected.

Our analysis showed that users tweet less frequently when they are protected, even for those users who stayed mostly protected ( $u_x$ ) during the data collection. It was also interesting to find that all user groups mention others more when they are public. However,  $u_o$  mention verified users or non-followers more while protected,  $u_x$  mentions them more while public, while there is no change for  $u_b$ . This indicates that users change their usual setting to mention verified users and non-followers.

We conducted two user surveys to answer the remaining research questions RQ3 and RQ4. We investigated the reasons behind the tweet visibility changes between public and protected to answer



Reasons to turn <b>public</b>	Public	Balanced	Protected
To reach a broader audience and get more interaction with my tweets	65.7%	50.4%	39.1%
To mention/reply to a user who does not follow me	49.5%	42.9%	41.3%
To retweet other users	49.5%	41.4%	31.5%
To gain more followers	43.4%	33.1%	27.2%
To quote tweet other users	33.3%	27.1%	21.7%
To enter to get giveaways or freebies	21.2%	21.8%	20.7%
To share articles or links	20.2%	20.3%	15.2%
To share pictures	23.2%	15%	16.3%
To associate a tweet with hashtags or trends publicly	24.2%	12.8%	14.1%
To boost the visibility, popularity, or ranking of a hashtag or topic	25.3%	13.5%	8.7%
To mention/reply to celebrities, famous people, or other VIPs	17.2%	9.0%	8.7%
To have a professional image	7.1%	14.3%	6.5%
To get customer service	9.1%	9.8%	6.5%
To boost the visibility of another user's tweet	11.1%	6.8%	8.7%
To find potential employment	0	5.3%	3.3%
To sell things or receive donations	4%	1.5%	0
Other	1.0%	1.5%	0
I did not change my tweet visibility settings to public before	2.0%	0.8%	7.6%

**Table 6: Reasons to turn public. Answers divided based on if their account is more often protected, public, or balanced. Percentages are out of the total number of mostly public, balanced, and mostly protected participants respectively.**

Reasons to turn <b>protected</b>	Public	Balanced	Protected
I wanted to prevent non-followers from seeing tweets with personal content	50.5%	52.6%	63.0%
People I know found my account and that made me uncomfortable	47.5%	51.9%	44.6%
To get a sense of privacy	35.4%	38.3%	46.7%
To prevent people I know, such as friends and family, from seeing my tweets	36.4%	38.3%	42.4%
To prevent interactions from strangers	30.3%	26.3%	38.0%
To avoid harassment	23.2%	27.1%	22.8%
To talk about a sensitive, controversial, or political topics freely	17.2%	15.8%	14.1%
To take a temporary break from interactions with non-followers	19.2%	14.3%	4.3%
I did not want people to retweet me	11.1%	8.3%	14.1%
To archive the account without deleting it	10.1%	9.8%	8.7%
To tweet about someone without them being able to see the tweets	9.1%	9.0%	8.7%
I did not want people to quote tweet me	13.1%	6.0%	7.6%
To share pictures	5.1%	6.8%	10.9%
My tweet unexpectedly went viral	6.1%	2.3%	6.5%
To share content that is not safe for work (NSFW)	5.1%	4.5%	3.3%
To quote tweet other users	3.0%	4.5%	1.1%
To prevent account suspension	2.0%	2.3%	5.4%
To share articles or links	4.0%	0.8%	3.3%
To retweet other users	2.0%	3.0%	0
Other	3.0%	1.5%	1.1%
I did not change my tweet visibility settings to protected before	1.0%	3.0%	0

**Table 7: Reasons to turn protected. Answers divided based on if their account is more often protected, public, or balanced. Percentages are out of the total number of mostly public, balanced, and mostly protected participants respectively.**

our RQ3. Over 40% of our participants changed their tweet visibility settings to public so they can mention non-followers. Other interaction-based reasons such as retweeting, quote tweeting, and entering giveaways were popular reasons to turn public. On the other hand, our participants changed their tweet visibility to protected so they can limit the audience and unwanted interactions. Our participants were also wary of people they know, e.g. family and friends, finding their accounts. Avoiding harassment and being able to talk freely about possibly controversial topics were among the most popular reasons to turn protected.

Our RQ4 concerns the strategies users utilize to manage their audience and interactions. The most popular strategy our participants employed was deleting tweets before moving the account from protected to public. Hidden preventions such as soft-blocking and muting were more popular compared to blocking, which can be noticed by the blocked person. Our participants also changed their account visibility to protected when they are unable to respond to interactions or deactivated their accounts so they can prevent all kinds of interactions including interactions from followers.

Control Visibility & Interactions	Public	Balanced	Protected
Delete some or all tweets when moving from protected to public	49.5%	57.9%	52.2%
Remove a follower without blocking them (soft blocking)	46.5%	43.6%	54.3%
Mute a follower so you can't see their interactions	44.4%	39.8%	39.1%
Block followers to prevent them from interacting even when your account is public	34.3%	36.8%	33.7%
Temporarily deactivate your account to prevent all interaction for a time	17.2%	16.5%	10.9%
Change to protected when not logged in or otherwise unable to respond to interactions	12.1%	10.5%	7.6%
Have a clear list of engagement rules prominently shown or linked to that detail what is acceptable interaction or following behaviour	4.0%	0.8%	1.1%
Other	0	0	1.1%
I have not used any of the above	6.1%	4.5%	3.3%

**Table 8: Actions taken to control who can see and interact with tweets. Answers divided based on if their account is more often protected, public, or balanced. Percentages are out of the total number of mostly public, balanced, and mostly protected participants respectively.**

## 5.2 Implications

Twitter’s relatively simplistic privacy settings imply that accounts are either public, where anyone on the internet can see all their tweets, or protected, where only followers can see the tweets. An observation from this work is that for many users the visibility of their tweets is not a static setting but instead one that changes as their needs and circumstances change. For these users, it is inaccurate to think of them as simply public or protected as they are changing their settings to get a mix of the affordances granted by both those states.

Current setting language also focuses around the protection of the tweets themselves rather than the person: “Unprotecting your Tweets will cause any previously protected Tweets to be made public” [7]. We find that users made changes to settings to protect themselves (e.g. harassment) or to create safe spaces (e.g. discussing controversial topics), where the focus in both cases was around the users themselves and their ability to have conversations with other users in a private space rather than to necessarily protect the tweets. The overlap here is obviously large since the conversations they are having are happening via tweets so they have to protect the tweets in order to create a safe space to have these conversations. But the conceptual difference is quite important to understand as it contextualizes the different actions users are taking.

Privacy expectations and notions of people are dynamic [3] representing a continual effort to control their presentation of self in a social context [16]. While Twitter’s binary setting option is simple, which helps with understanding it, that simplicity also makes it more challenging to conduct fine grain boundary management. Particularly since users are not just moving themselves between a protected and public space, the way they might when they leave the house to go to a coffee shop, they are also bringing all their past tweets with them when they move between audience spaces, closer to holding an open house where anyone can enter the previously private space. The bringing of past events with them when they move between audience spaces makes proper boundary management more challenging since the whole tweet feed has to be curated to be appropriate for the new audience which can be cumbersome [21, 54, 57]. Alternatively users may rely on self-censorship [35, 45] where they refrain from sharing sensitive information even when protected to limit the risks of it becoming public later. Approaches like these, or even lack of awareness, may

be why 46% of survey respondents indicated that they don’t delete tweets when changing from protected to public. Another reason for not deleting past tweets might be the archival value of historical tweets and protecting meaningful self-representation [21, 54, 57].

Some of our users definitely curated their feeds when switching between protected and public. Deleting tweets was the most common audience control action selected in the survey, and in the tweet dataset we noticed fewer tweets when people were protected than when they were public which *might* be an indication that users are deleting tweets before switching. Participants also mentioned the need to briefly protect a specific tweet or be able to briefly engage with a public account. For example, a participant mentioned being protected to share a private Instagram account with their followers to gain “*a momentary bit of privacy before of deleting the tweet and making my account public again.*” Prior work has also shown that users delete their tweets for various reasons including preventing harassment, revealing too much information [46], or simple typos [2]. In line with Mondal et al. [37], we find that users utilize deletion to protect their privacy, control information flow, and manage audiences.

Users both want to “achieve what the platform could offer to its fullest” while also “keep tweeting freely”. They are not just looking for a way to create a safe space, they also want to be able to interact with the wider community by doing things like replying to a public tweet or making parts of their tweet stream public. Hence, we see the active management of tweet visibility in our dataset where we find that some users changed their settings more than 200 times in three months. The nature of the Twitter usage also seems to affect this behaviour. As a participant stated: “*how much you change your public settings depends on which side of twitter you are. I have a fan account so I get more harassment than if I had a personal account and I change my setting more often.*”

One solution our participants made use of was to have multiple accounts which had different settings, similar to Stutzman and Hartzog [48]. A participant stated: “*in my twitter bubble it is very common to have two accounts, one public (and anonymous) and one private for the few close and trusted friends made on the platform, talking about private matters and sharing images etc.*” Instagram users use “finstas” for similar purposes [56]. Some users even appear to share their protected accounts on their bio and ask to be “private moots (mutual follow)” where they usually add only protected users

to their network to minimize information leak, which can happen if they have conversations with public accounts [25, 26].

### 5.3 Design Recommendations

The largest design recommendation we have is that Twitter should think about tweet visibility in terms of conversations, time frames, or audiences rather than focusing on the account level only.

*Visibility Pinning of Tweets:* Provide users with the ability to “pin” a tweet as either public or protected so that it stays that way even when the account changes visibility. Doing so would allow users to more easily share a single tweet publicly or privately without needing to change their whole account visibility. It would also allow predominately protected users to share a public set of tweets that would allow them to attract new followers without opening their entire account up.

*Paired Accounts:* Directly support the practice of having both public and protected accounts. Users already use this approach to manage the visibility of their tweets and their audiences, but they are doing it ad hoc. Twitter could more directly support this practice by offering to create two linked accounts and making it easier to switch between them to post to appropriate audiences. Doing so would allow for better boundary management as well as assist users who are looking to gain followers by having some tweets public.

*Temporal Settings:* Make some settings time based so that they encompass a time period or a conversation stream rather than just a single tweet. A user may change to protected to have a private conversation with their followers and then change back to public, but then they would need to go and delete all the tweets happening in that time frame. It would be helpful to users to have the ability to delete tweets that happened within a frame of time. It may also be helpful to have tweets that disappear after a set time frame or disappear when the account visibility changes. Posts with time-limits that disappear after a set time is already adopted by various social media platforms including Snapchat, Instagram, and WeChat. These features help users to cope with temporal context collapse to a degree and users appreciate these ephemeral posts [21]. However, the time-limit should be controlled by the users instead of the platform. According to Yilmaz et al. [57], users found the automated deletion is beneficial only when the users are in control of the deletion date. Tweet deletion is also a common practice among users who do not necessarily switch their account visibility [2, 37, 46] so this feature may benefit them too.

*Limiting Interactions:* As there were participants who made their account public to interact, there were also participants who protected their accounts to limit interactions from non-followers. Some of our participants even deactivated their accounts temporarily to prevent all interactions for a while. There were also some participants who chose to limit interactions when they are unable to respond to them. Twitter recently introduced a feature for users to limit the replies to specific tweets. However, other interactions such as retweeting, quote tweeting, and liking are still possible for these tweets. Especially, quote tweets can be used as replies by other users when replies are limited, which also increases the

reach of the original tweet more than the replies [15]. Users can stop all interaction by deactivating their accounts up to 30 days without deletion but deactivation will hide users’ tweets, even from the user themselves. Enabling users to stop all interactions without changing their account visibility to protected or deactivating can help the participants adopting those strategies to limit/stop interactions. Additionally, prior work has shown that people see an archival value in their social media accounts [21, 54, 58]. This feature will benefit nearly 10% of our participants who also wanted to archive their accounts without deleting it.

*Removing Followers:* Allow users to remove followers directly rather than requiring them to soft-block them. Doing so would allow users to more easily curate their followers when they switch between public and protected. It would also be important to consider allowing them to do so silently without raising a notification to the other person since people prefer online sanctions that can create “plausible deniability” rather than visible ones, especially when dealing with followers who are strong-ties [44].<sup>4</sup>

### 5.4 Limitations

We curated our initial set of users by collecting mentioned protected accounts. Hence, our dataset is skewed towards users who stay protected most of the time. However, we observed that the users we collected tweets from represent a wide range of percentage of time spent protected vs private as well as exhibiting a good distribution over that range. Another limitation is the representation of mostly protected users in  $t_x$  and  $t_o$ . We were only able to collect tweets of accounts when they are public, so in the case where a user deletes their tweets before changing their tweets to public, we would not be able to collect the deleted tweets. Hence, for some users we might be getting only the  $t_x$  tweets that users are comfortable sharing publicly.

Our survey participants also represent a younger population, who may have a different view on privacy and have different goals than people of other age groups. In the United States the median Twitter user age is 40 and 62% of worldwide Twitter users are younger than 34 [10, 55]. Which means that our survey participants are indeed younger than a typical Twitter user but not greatly so. We recruited our survey participants from the crowdsourcing platform Prolific Academic and crowdsourcing participants tend to be more privacy-conscious than the general population [24]. This population source may be the cause of the higher percentage of mostly protected accounts in our participants (28%) compared to their general percentage among Twitter users (5%) [31]. Self reporting is also known to have biases, particularly around memory [39], and our survey relies on answering questions about past events.

The first survey was also intended to create an initial list of reasons that people switch between public and protected rather than create a comprehensive list. The number of participants surveyed, in our view, is high enough to identify common issues, but too few to really understand the scope of all possible reasons for switching. Also, as noted above, our participant sample is younger than an average Twitter user which also likely impacted the types of reasons they provided. We try and balance this issue by also including

<sup>4</sup>Twitter recently provided a feature for Twitter Web users to remove followers. However, this feature was introduced after we conducted our studies.

reasons found in prior work, but it is quite likely that if we were to more aggressively survey people from different age ranges, cultures, physical locations, or other demographics we would likely find a wider set of reasons for switching.

## 6 CONCLUSION

In this study, we investigated Twitter users' privacy setting switching behaviour and the reasons behind the changes. To do so, we collected the account visibilities of 107K initially protected users for three months and found that nearly 40% of them changed their settings at least once. We also collected the switchers' tweets to understand whether their sharing behaviour differs when they are protected vs public. We find that users utilize the privacy settings dynamically, sometimes changing as much as daily. They send tweets with mentions and hashtags more when they are public compared to protected. We coupled our Twitter data with two user surveys to get insight into the potential reasons behind the changes. We find that users change their accounts to protected to control their audience and interactions. On the other hand, users prefer to change their account visibility to public to use Twitter features freely, possibly at the expense of their privacy. We also suggest some design implications to protect the privacy of the users while enabling them to experience what the platform offers fully.

## ACKNOWLEDGMENTS

This work was supported in part by the EPSRC DTA award, funded by the UK Engineering and Physical Sciences Research Council and the University of Edinburgh. We thank everyone associated with the TULiPS Lab and SMASH Group at the University of Edinburgh for helpful discussions and feedback.

## REFERENCES

- [1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2020. Secrets and likes: the drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology* 30, 4 (2020), 736–758.
- [2] Hazim Almuhiemedi, Shomir Wilson, Bin Liu, Norman Sadeh, and Alessandro Acquisti. 2013. Tweets are forever: a large-scale quantitative analysis of deleted tweets. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 897–908.
- [3] Irwin Altman. 1976. A conceptual analysis. *Environment and behavior* 8, 1 (1976), 7–29.
- [4] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 3906–3918.
- [5] Eric PS Baumer, Phil Adams, Vera D Khovanskaya, Tony C Liao, Madeline E Smith, Victoria Schwanda Sosik, and Kaiton Williams. 2013. Limiting, leaving, and (re) lapsing: an exploration of facebook non-use practices and experiences. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 3257–3266.
- [6] Eric PS Baumer, Rui Sun, and Peter Schaedler. 2018. Departing and Returning: Sense of Agency As an Organizing Concept for Understanding Social Media Non/Use Transitions. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–19.
- [7] Twitter Help Center. 2021. About public and protected tweets. Retrieved April 15, 2021 from <https://help.twitter.com/en/safety-and-security/public-and-protected-tweets>
- [8] Yoon Hyung Choi and Natalya N Bazarova. 2015. Self-disclosure characteristics and motivations in social media: Extending the functional model to multiple social network sites. *Human Communication Research* 41, 4 (2015), 480–500.
- [9] Sauvik Das and Adam Kramer. 2013. Self-censorship on Facebook. In *Seventh international AAI conference on weblogs and social media*.
- [10] Statista Research Department. 2021. Distribution of Twitter users worldwide as of July 2021, by age. Retrieved Sep 4, 2021 from <https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/>
- [11] Brianna Dym and Casey Fiesler. 2018. Vulnerable and online: Fandom's case for stronger privacy norms and tools. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 329–332.
- [12] Nicole B Ellison, Jessica Vitak, Charles Steinfield, Rebecca Gray, and Cliff Lampe. 2011. Negotiating privacy concerns and social capital needs in a social media environment. In *Privacy online*. Springer, 19–32.
- [13] Casey Fiesler, Michaelanne Dye, Jessica L Feuston, Chaya Hiruncharoenavate, Clayton J Hutto, Shannon Morrison, Parisa Khanipour Roshan, Umashanthi Pavalanathan, Amy S Bruckman, Munmun De Choudhury, et al. 2017. What (or who) is public?: Privacy settings and social media content sharing. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 567–580.
- [14] Casey Fiesler and Nicholas Proferes. 2018. "Participant" perceptions of Twitter research ethics. *Social Media+ Society* 4, 1 (2018), 2056305118763366.
- [15] Kiran Garimella, Ingmar Weber, and Munmun De Choudhury. 2016. Quote rts on twitter: usage of the new feature for political discourse. In *Proceedings of the 8th ACM Conference on Web Science*. 200–204.
- [16] Erving Goffman. 1959. *The presentation of self in everyday life*. Garden City, NY Double Day.
- [17] Sukeshini A Grandhi, Linda Plotnick, and Starr Roxanne Hiltz. 2019. Do I Stay or Do I Go? Motivations and Decision Making in Social Media Non-use and Reversion. *Proceedings of the ACM on Human-Computer Interaction* 3, GROUP (2019), 1–27.
- [18] Ralph Gross and Alessandro Acquisti. 2005. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*. ACM, 71–80.
- [19] Karen Holtzblatt and Hugh Beyer. 2017. 6 - The Affinity Diagram. In *Contextual Design (Second Edition)*, Karen Holtzblatt and Hugh Beyer (Eds.). Morgan Kaufmann, Boston, 127–146. <https://doi.org/10.1016/B978-0-12-800894-2.00006-5>
- [20] Roberto Hoyle, Srijita Das, Apu Kapadia, Adam J. Lee, and Kami Vaniea. 2017. Viewing the Viewers: Publishers' Desires and Viewers' Privacy Concerns in Social Networks. In *Proceedings of the Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*. 555–566.
- [21] Xiaoyun Huang, Jessica Vitak, and Yla Tausczik. 2020. "You Don't Have To Know My Past": How WeChat Moments Users Manage Their Evolving Self-Presentation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [22] Shagun Jhaver, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Online harassment and content moderation: The case of blocklists. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 2 (2018), 1–33.
- [23] Maritza Johnson, Serge Egelman, and Steven M Bellovin. 2012. Facebook and privacy: it's complicated. In *Proceedings of the eighth symposium on usable privacy and security*. 1–15.
- [24] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. 2014. Privacy Attitudes of Mechanical Turk Workers and the U.S. Public. In *Symposium On Usable Privacy and Security (SOUPS '14)*. 37–49.
- [25] Dilara Keküllüoğlu, Walid Magdy, and Kami Vaniea. 2020. Analysing privacy leakage of life events on Twitter. In *Proceedings of the 12th ACM Conference on Web Science*.
- [26] Dilara Keküllüoğlu, Walid Magdy, and Kami Vaniea. 2022. From an Authentication Question to a Public Social Event: Characterizing Birthday Sharing on Twitter. In *Proceedings of The 16th International AAI Conference on Weblogs and Social Media (ICWSM'22)*.
- [27] Cliff Lampe, Jessica Vitak, and Nicole Ellison. 2013. Users and nonusers: Interactions between levels of adoption and social capital. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 809–820.
- [28] Airo Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. 2011. We're in it together: interpersonal management of disclosure in social network services. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 3217–3226.
- [29] Hai Liang, Fei Shen, and King-wa Fu. 2017. Privacy protection and self-disclosure across societies: A study of global Twitter users. *new media & society* 19, 9 (2017), 1476–1497.
- [30] Yabing Liu, Krishna P Gummadi, Balachander Krishnamurthy, and Alan Mislove. 2011. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. 61–70.
- [31] Yabing Liu, Chloe Kliman-Silver, and Alan Mislove. 2014. The tweets they are a-changin': Evolution of Twitter users and behavior. In *Eighth International AAI Conference on Weblogs and Social Media*.
- [32] Michelle Madejski, Maritza Johnson, and Steven M Bellovin. 2012. A study of privacy settings errors in an online social network. In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*. IEEE, 340–345.
- [33] Michelle Madejski, Maritza Lupe Johnson, and Steven Michael Bellovin. 2011. The failure of online social network privacy settings. (2011).
- [34] Merja Mahrt, Katrin Weller, and Isabella Peters. 2014. Twitter in scholarly communication. *Twitter and society* 89 (2014), 399–410.

- [35] Alice E Marwick and Danah Boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society* 13, 1 (2011), 114–133.
- [36] Alice E Marwick and Danah Boyd. 2014. Networked privacy: How teenagers negotiate context in social media. *New media & society* 16, 7 (2014), 1051–1067.
- [37] Mainack Mondal, Johnnatan Messias, Saptarshi Ghosh, Krishna P Gummadi, and Aniket Kate. 2016. Forgetting in social media: Understanding and controlling longitudinal exposure of socially shared data. In *Twelfth Symposium on Usable Privacy and Security (SOUPS) 2016*. 287–299.
- [38] Mainack Mondal, Günce Su Yilmaz, Noah Hirsch, Mohammad Taha Khan, Michael Tang, Christopher Tran, Chris Kanich, Blase Ur, and Elena Zheleva. 2019. Moving beyond set-it-and-forget-it privacy settings on social media. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. 991–1008.
- [39] Andreas Möller, Matthias Kranz, Barbara Schmid, Luis Roalter, and Stefan Diewald. 2013. Investigating self-reporting behavior in long-term studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Paris, France) (CHI '13)*. 2931–2940.
- [40] Yee Man Margaret Ng. 2020. Re-examining the innovation post-adoption process: the case of Twitter discontinuance. *Computers in Human Behavior* 103 (2020), 48–56.
- [41] Helen Nissenbaum. 2009. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.
- [42] Sandra Petronio. 2015. Communication privacy management theory. *The International Encyclopedia of Interpersonal Communication* (2015), 1–9.
- [43] Prolific Academic 2019. Prolific Academic. Retrieved September 15, 2019 from <https://www.prolific.co/>
- [44] Yasmeen Rashidi, Apu Kapadia, Christena Nippert-Eng, and Norman Makoto Su. 2020. "It's easier than causing confrontation": Sanctioning Strategies to Maintain Social Norms and Privacy on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–25.
- [45] Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. 2013. The post that wasn't: exploring self-censorship on Facebook. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 793–802.
- [46] Manya Sleeper, Justin Cranshaw, Patrick Gage Kelley, Blase Ur, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. 2013. "I read my Twitter the next morning and was astonished" a conversational perspective on Twitter regrets. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3277–3286.
- [47] Katherine Strater and Heather Richter Lipford. 2008. Strategies and struggles with privacy in an online social networking community. *People and Computers XXII Culture, Creativity, Interaction 22* (2008), 111–119.
- [48] Frederic Stutzman and Woodrow Hartzog. 2012. Boundary regulation in social media. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 769–778.
- [49] Sue Yeon Syn and Sanghee Oh. 2015. Why do social network site users share information on Facebook and Twitter? *Journal of Information Science* 41, 5 (2015), 553–569.
- [50] Brendan Van Alsenoy, Valerie Verdoodt, Rob Heyman, Ellen Wauters, Jef Ausloos, and Günes Acar. 2015. From social media service to advertising network: a critical analysis of Facebook's Revised Policies and Terms. (2015).
- [51] George Veletsianos, Shandell Houlden, Jaigris Hodson, and Chandell Gosse. 2018. Women scholars' experiences with online harassment and abuse: Self-protection, resistance, acceptance, and self-blame. *New Media & Society* 20, 12 (2018), 4689–4708.
- [52] Jessica Vitak. 2012. The impact of context collapse and privacy on social network site disclosures. *Journal of broadcasting & electronic media* 56, 4 (2012), 451–470.
- [53] Jessica Vitak, Stacy Blasiola, Sameer Patil, and Eden Litt. 2015. Balancing audience and privacy tensions on social network sites: Strategies of highly engaged users. *International Journal of Communication* 9 (2015), 20.
- [54] Jessica Vitak and Jinyoung Kim. 2014. "You can't block people offline" examining how facebook's affordances shape the disclosure process. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 461–474.
- [55] Stefan Wojcik and Adam Hughes. 2019. Sizing up Twitter users. *Washington, DC: Pew Research Center* (2019). <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>
- [56] Sijia Xiao, Danaë Metaxa, Joon Sung Park, Karrie Karahalios, and Niloufar Salehi. 2020. Random, Messy, Funny, Raw: Finstas as Intimate Reconfigurations of Social Media. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [57] Günce Su Yilmaz, Fiona Gasaway, Blase Ur, and Mainack Mondal. 2021. Perceptions of Retrospective Edits, Changes, and Deletion on Social Media. In *Proceedings of the Fifteenth International AAAI Conference on Web and Social Media (ICWSM'21)*.
- [58] Xuan Zhao, Niloufar Salehi, Sasha Naranjit, Sara Alwaalan, Stephen Voida, and Dan Cosley. 2013. The many faces of Facebook: Experiencing social media as performance, exhibition, and personal archive. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1–10.

## A CURATION OF REASONS FOR SWITCHING ACCOUNT VISIBILITY

Reasons to turn <b>public</b>	Twitter Data	Free-Text Survey
To reach a broader audience and get more interaction with my tweets		+
To gain more followers		+
To mention/reply to a user who does not follow me	+	+
To find potential employment		
To have a professional image		+
To sell things or receive donations		
To enter to get giveaways or freebies		+
To retweet other users	+	+
To quote tweet other users	+	+
To associate a tweet with hashtags or trends publicly	+	+
To boost the visibility, popularity, or ranking of a hashtag or topic		+
To boost the visibility of another user's tweet		
To share articles or links	+	
To share pictures	+	+
To get customer service		
To mention/reply to celebrities, famous people, or other VIPs	+	+

**Table 9: Reasons to turn public - Options given in the survey and their sources**

Reasons to turn <b>protected</b>	Twitter Data	Free-Text Survey
People I know found my account and that made me uncomfortable		+
My tweet unexpectedly went viral		+
I wanted to prevent non-followers from seeing tweets with personal content		+
To prevent people I know, such as friends and family, from seeing my tweets		+
To archive the account without deleting it		+
To avoid harassment		+
To tweet about someone without them being able to see the tweets		+
To prevent account suspension		+
I did not want people to retweet me	+	
I did not want people to quote tweet me	+	
To talk about a sensitive, controversial, or political topics freely		+
To prevent interactions from strangers		+
To share pictures	+	+
To share content that is not safe for work (NSFW)		+
To retweet other users	+	
To quote tweet other users	+	+
To get a sense of privacy		+
To share articles or links	+	
To take a temporary break from interactions with non-followers		+

**Table 10: Reasons to turn protected - Options given in the survey and their sources**