A Study of Users' Privacy Preferences for Data Sharing on Symptoms-Tracking/Health App

Hafiz Asif
Rutgers University, New Jersey
USA

Jaideep Vaidya
Rutgers University, New Jersey
USA

ABSTRACT

Symptoms-tracking applications allow crowdsensing of health and location related data from individuals to track the spread and outbreaks of infectious diseases. During the COVID-19 pandemic, for the first time in history, these apps were widely adopted across the world to combat the pandemic. However, due to the sensitive nature of the data collected by these apps, serious privacy concerns were raised and apps were critiqued for their insufficient privacy safeguards. The Covid Nearby project was launched to develop a privacy-focused symptoms-tracking app and to understand the privacy preferences of users in health emergencies.

In this work, we draw on the insights from the Covid Nearby users' data, and present an analysis of the significantly varying trends in users' privacy preferences with respect to demographics, attitude towards information sharing, and health concerns, e.g. after being possibly exposed to COVID-19. These results and insights can inform health informatics researchers and policy designers in developing more socially acceptable health apps in the future.

CCS CONCEPTS

• **Security and privacy → Privacy protections**; **Usability in security and privacy**.

KEYWORDS

privacy preferences; COVID-19; symptoms-tracking; user study

ACM Reference Format:
Hafiz Asif and Jaideep Vaidya. 2022. A Study of Users' Privacy Preferences for Data Sharing on Symptoms-Tracking/Health App. In *Proceedings of the 2022 Workshop on Privacy in the Electronic Society (WPES '22), November 7, 2022, Los Angeles, CA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3559613.3563202

1 INTRODUCTION

Since the outset of the Corona Virus Disease 2019 (COVID-19) pandemic, a vast number of tracking and tracing phone and web applications (apps) have been developed across the world to curb the spread of COVID-19. Since these apps work by gathering health and location related sensitive information from the individuals, immediately, there was an uproar about the infringements on freedom and privacy rights. This led to many research studies, from analyzing app's privacy [27] to peoples preferences and expectation of privacy and utility from these apps [1, 15, 24–26, 28].

In contrast to the existing work, most of which are based on focus group and surveys, here we analyze the privacy preferences of the users of a privacy-preserving symptoms-tracking app, Covid Nearby [2, 22]. While reporting the tracking data, the app provides the provable guarantee of differential privacy [21] to protect users data and their whereabouts. The assurance and protection of privacy was at the center of the design, development and publicity campaign of Covid Nearby. Thus, the study adds a unique perspective to the existing literature.

The symptoms-tracking apps are unique in the sense that on one hand, they help user assess their and their community's risk amid a health emergency (e.g. tracking new outbreaks and spread of the disease). On the other hand, the app is only effective if the users share their data that is personal and private. Thus, to use these apps users rely on their perceived privacy utility trade-offs [24].

The aim of the study was to elucidate the trends and preference in users' privacy attitude on health apps, especially, on a symptoms-tracking app during a global pandemic, a health emergency.

Data collection and study design: We made the app available via web, Android, and iOS platforms and publicised them via presentation at various forums as well as via news. The study was approved by the institutional review board. When using the app, via a short questionnaire, the users were asked to share their COVID-19 health-related information (e.g. self-reported symptoms and possible exposure to COVID-19), demographic data (e.g. age, gender, and location), and a privacy level for data protection. This questionnaire was designed in a collaboration with epidemiology, data privacy, and behavioral informatics experts.

For the 'Gender' feature, the users could report their self-described gender identity. Given the sensitivity of this feature, we also provided an option to not disclose this information. This non-disclosure option was also provided for 'Tested' feature, , i.e. whether the user was tested for COVID-19. We did this because COVID-19 (at least in the first year) carried enormous stigma, e.g. a woman who had tested positive for COVID-19 could not be located since she had given fake information [4]. We will see in Section 2 how the option of non-disclosure can help characterize users' privacy preference.

To gain insights into users' privacy preferences, we asked them to select a level of privacy at the end of the questionnaire, just before submitting their data (as shown in Appendix, Fig. 5). The user could increase or decrease the privacy (by two levels) from the

WPES '22, November 7, 2022, Los Angeles, CA, USA.

ACM ISBN 978-1-4503-9873-2/22/11…$15.00
https://doi.org/10.1145/3559613.3563202

default[1]. Importantly, note that the users were asked to make this choice in the context of the privacy-utility trade-off of their data.

Additionally, we aimed to identify the *vigilant users*—who are vigilant about their privacy and the privacy policies of the app—and understand their privacy preferences compared to other users, referred to as *casual users*. To accomplish this, we noticed whether a user opted to read the privacy-related information [3] (by clicking on the provided link as shown in Fig. 5).

Data Description and Users' Demographic: Table 1 gives the features of the data we consider in this study. The data consists of 978 users ($N = 978$)[2] from 41 countries, but the United States (US) users constituted the supermajority, 773 (78%). Among other countries, a significant participation was from the United Kingdom (UK), Australia (AU), and Canada (CA) (see Fig. 1(a)). The median age (*MA*) group of the users from US and AU was 35-44 (years), while for the rest (i.e. UK, CA, and OT=others), it was smaller: 25-34.

In terms of gender, males (49.0%) and females (48.0%) had roughly equal participation with female users being younger (*MA* =25-34) than male users (*MA* =35-44). The participation, however, across different age groups was different: From the younger age group (18-44), more females participated while more males participated from the older age group (45 and older)—see Fig. 1(c). The Figure also shows that the participation reduces in higher age groups.

About 3% of the users (*MA* =35-44) did not disclose their gender identity and only 5 users identified as gender non-binary.

The US users had median age of 35-44 and were from 44 different states. The majority users were from New Jersey (NJ). Among other states, a significant participation was from New York (NY), California (CA), Florida (FL), and Texas (TX) (see Fig. 1(b) for details). The users from NJ were the youngest (*MA* =25-34), followed by CA, NY, and TX (*MA* =35-44), and FL (*MA* =45-54). The relative number and youth of NJ users is not surprising since the study

[1]Covid Nearby used differential privacy to protect data, where the higher level of privacy corresponds to a stronger privacy, see [2] for details. We note that we provided everyone the same default privacy level, and used the question to assess the privacy-utility trade-offs.

[2]The IRB approval was for 18+; so we removed 140 users who reported age < 18.

Table 1: Data features' description.

Feature	Description and Values
Age	User's age group *18-21, 22-24, 25-34, 35-44, 45-54, 55-64, >= 65*
Gender	User's gender, with an optional non-disclosure *Male, Female, Non-Binary, Undisclosed*
User-type	Whether a user is vigilant regarding privacy *vigilant user, casual user*
Location	Location coordinates, converted to: *country and state*
Privacy	Data protection level selected by the user *-2* (Min), *-1* (Low), *0* (Default), *1* (High), *2* (Max)
Tested	Whether the user was tested for COVID-19 *Yes, No, Undisclosed*
Symptomatic	The user is experiencing COVID-19 symptoms *Yes, No*
Exposed	The user is possibly exposed to COVID-19 *Yes, No*

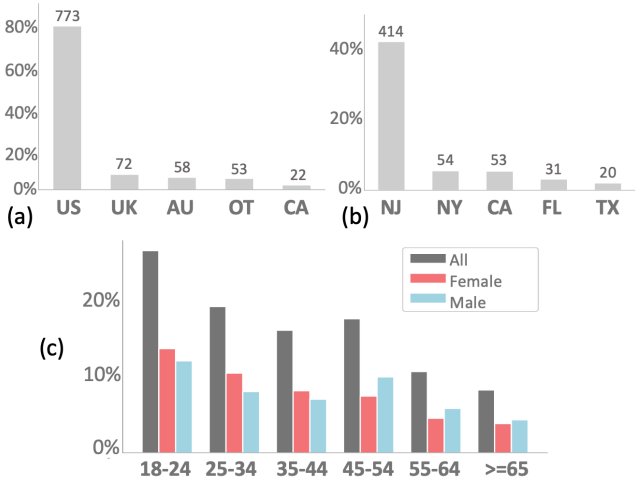

Figure 1: Users' Demographics. The plots (a), (b), and (c) respectively give the participants per country (OT stands for others, countries not mentioned by name), US states (with most users), and age group. Non-binary and undisclosed gender were excluded from (c) due to insufficient data.

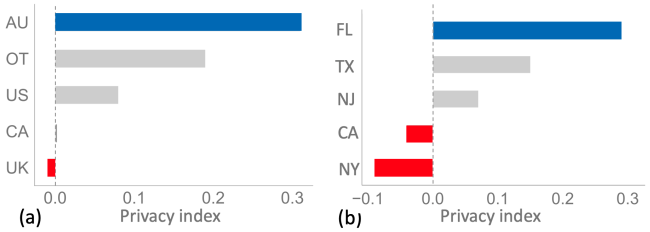

Figure 2: Privacy comparison across countries and the US states. (a) is for countries. (b) is for 5 US states with most users.

was conducted at Rutgers, the State University of NJ, and was also initially publicized within the university.

Finally, we note that 9% of the users reported being possibly exposed to COVID-19 in the past 14 days, and 12% reported feeling COVID-19 related symptoms. 30% reported getting a test for COVID-19 while about 4% of the user did not disclose this information.

2 ANALYSIS: TRENDS AND INSIGHTS

To analyze users' privacy preferences and trends at a group level, we will use *privacy index*, abbreviated as *PI*. *Privacy index* (*PI*) is the average of privacy levels (ranging from −2 to +2, see Table 1) of a given group (e.g. the users from the US); and it measures how low or high the group regards its privacy. Thus, privacy index value ranges from −2, for the lowest privacy, to +2, for the highest privacy (recall from Table 1 that the default level is assigned the value 0).

The privacy index of the data over all the users was 0.1 despite 81% of the users accepting the default privacy. This indicates that overall the users preferred a higher privacy than the default level; however, when the users change their privacy, more often they opt for higher privacy—in fact, a significant proportion (about 8%) increase the privacy to the Max level. However, the index as well as the privacy-adjustment behavior differs by age groups—especially, in the youngest and the oldest users—and by location as we discuss below in detail.

Privacy by location: We first look at the privacy preferences per country and how they compare with each other, given in Fig. 2(a). AU ($PI = 0.3$) has the highest privacy index, while the UK ($PI = -0.014$) has the lowest and is much closer to that of CA ($PI = 0$). The US (0.08) was third in term of privacy index.

To see the users' predilections towards higher or lower privacy, we look at the percentage of users (from each country) who changed the default privacy level. AU users changed the default the most (22%), where 19% consisted of the users who changed the default to increase the privacy level. The second was the US: 19% changed, wherein 10% increased the privacy level. 17% of the OT users changed default privacy, wherein 13% increased it. Although 17% of UK users changed the default privacy, only 6% increased it.

We note that the user per each country in OT group is extremely small; so the general applicability of the index comparison may be not well supported. Similar is the case with CA: none of the CA users changed the default. This may be due to a biased and small sample (22 users $\approx$ 2%, see Fig. 1(a)).

Next, let us look into the US (see Fig. 2(b)), focusing on the top-5 states (in terms of participation). FL ($PI = 0.29$) users opted for higher privacy compared to the other states. NY ($PI = -0.09$) and CA ($PI = -0.04$) had the smallest privacy index–both were negative though very close to 0. NJ ($PI = 0.07$) had the third smallest (or largest) privacy index, and in contrast to NY and CA, it was positive. Furthermore, as per PI, NY and CA are similar to each other; on the other hand, TX ($PI = 0.15$) and FL were similar to each other, and had a higher privacy index. NJ stands in the middle though it is closer to TX than its geographical neighbor NY.

Privacy by feature: We now look at how privacy preferences are related to different features in the data. For instance, one may suspect that the person who is possibly exposed to COVID-19 or experiencing COVID-19 symptoms will value the utility of the data a bit higher and privacy a little lower. Data does indicate this—see Fig. 3. In addition, there was a medium level correlation—measured by Matthews correlation coefficient (MCC) [20]—between a user being symptomatic and possibly exposed to COVID-19: MCC = 0.3. However, the changing or increasing privacy was not significantly correlated with being symptomatic or possibly exposed to COVID-19 (MCC ≈ -0.04 for both).

The users with highest privacy level (as per PI) were who chose to not disclose their gender or whether they received a COVID-19 test. Not disclosing 'Gender' or if one was 'Tested' was correlated with opting for higher privacy (χ^2-test was significant with p-value 0.0025 for significance level of 0.05 and MCC $\approx$ 0.14) with an odds ratio (OR) = 2.9. Namely, if a user did not disclose some of the information, the user was three times more likely to increase the privacy level from default than the user who disclosed all the information.

Although females have a slightly higher PI than males, male users were twice as likely to be vigilant with OR = 1.9. Gender Non-binary users, on the other hand, had $PI = -0.01$—this, however, is not well supported due to the tiny sample size of 5 (we only include this observation for completeness).

Furthermore, vigilant users make up 4% of the total users. Most of the vigilant users were from the US, making up approximately

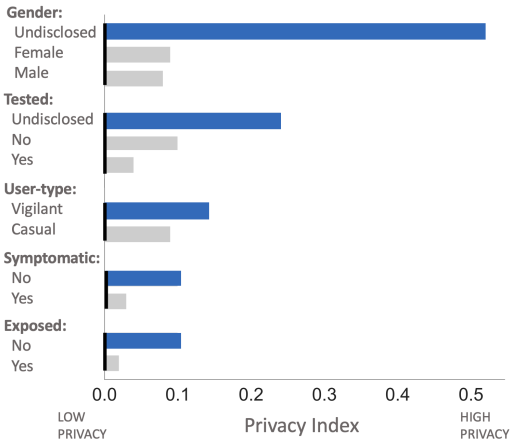

Figure 3: Privacy comparison across features.

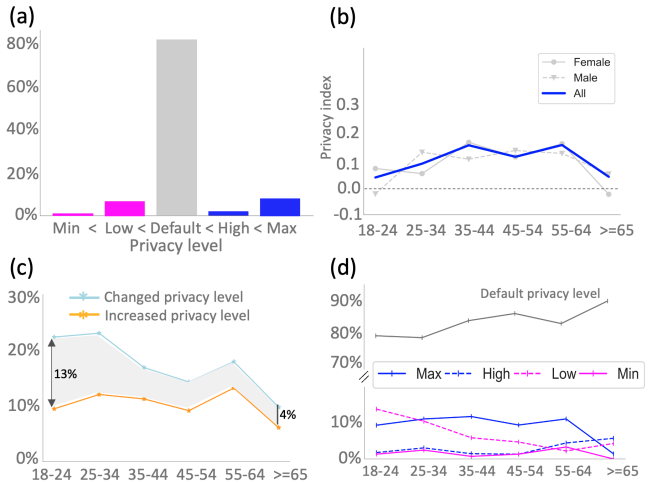

Figure 4: Privacy comparison across age groups. (a) shows the percentage users per each privacy level. (b) shows the privacy index per age group for all users, males, and females. (c) shows the percentage users per age group who changed or increased the privacy level; the narrowing gap between the curves indicates that more users change the default to increase the privacy. (d) shows the percentage of users per each privacy level within each age group; the percentages sum up to 100 for each age group.

80% of the vigilant users. Compared to casual users, they are more likely to change the default privacy level (MCC = 0.13 and p-value for χ^2-test = 0.04), and as per PI and the odds ratio (2.5), they are about twice as likely to opt for higher privacy.

Privacy by age: Across all age groups, 80% or more users did not change the default privacy level (Fig. 4(a)). Of the 19% users who change the default, 10% increased privacy though there is difference based on their age (discussed shortly). Furthermore, when the users increased their privacy, they (about 8%) increased it to the max level; and when they decreased it, they did not opt for the minimum level—only 1.4% of the users chose the Min level.

With increasing age, we observed a general trend in predilection towards higher privacy (see Fig. 4(b)) except for the oldest users, i.e. age group of 65 and older (>=65). Although the youngest (18-24) and the oldest users are closer in terms of their privacy index

with *PI*'s 0.034 and 0.037 respectively, they distinctly differ in their preferences, see Fig. 4(b)-(c).

The oldest users increased the privacy level much more often, about 60% of the times, while the youngest users often lower the privacy, about 60% of the times (this behavior can also be seen in Fig. 4(b),(c)). Another noticeable difference was that the *PI*'s for the males in youngest users and the females in the oldest users were negative, deviated significantly from the overall *PI* of their respective age groups, and were lower than the other gender (female and males respectively) in the same age group.

The mid age group 35-64 opt for higher privacy, compared to the rest—and this relationship is strong: χ^2-test on changing privacy level over age groups 18-34, 35-64, and >=65 was very significant with p-value=0.004. As the age increases, the users more often change the privacy level to increase it (see Fig. 4(b)). However, as the age increases, fewer users change the default.

Finally, we note from Fig. 4(d) that for ages 18-64, Max privacy level is chosen by approximately the same percentage of users. Further, as the age increases, the Low privacy level is chosen by fewer and fewer users expect for the age >= 65, and default is chosen by more (as mentioned earlier). The age group >=65 changes the default the least, but opts for Low (-1) and High (+1) privacy level more often than either extreme (i.e. Min or Max) and is unique in this sense. The only other group that opted for Low privacy level more often was 18-24 and is also unique in this sense.

3 DISCUSSION AND CONCLUSIONS

The adoption of health-related technology and smartphone apps is higher in younger individuals as indicated by our data and other studies [6]. In future, we believe such apps will be widely used in general as well as in health emergencies. Hence, it is imperative to understand the users' privacy preferences and take them into account in the design and development of health apps.

We must note that this study comes with its limitations and caveats that may restrict its wider and general applicability. Since much of the publicity was done online and via the university, this can lead to a biased sample. In addition, for certain groups, the data may be not sufficient. Since it is a user study, there is no control group or base line to compare the results (below we do place these results in the context of the previous work). However, this study points to various crucial and emerging trends that must be thoroughly investigated.

We observed a higher participation of females than males, especially, in age group (18-44), which is in contrast with a previous study [6]. However, that study focused on activity trackers and diet apps, and thus, the participants' health concerns were less severe than during the pandemic. Additionally, females opted for a higher privacy level than males. This can informs the developers to integrate higher privacy safeguards in female-specific apps (e.g. menstrual cycle tracking apps).

An overwhelming majority of users (81%) accepted the default privacy level, reaffirming the strong influence of the default [10, 11, 13, 17, 18]. However, this acceptability rate was much higher. For instance, in an online survey study about sharing private information, 54% participants accepted the default privacy. In another study, 56% of the users (age 18-29) chose the default privacy on online social networks; in contrast, we found the acceptability to be 77% in the similar age group.

The higher acceptability, we believe, is due to two main factor. One, our privacy-focused design and publicity campaign: we provided an endowed privacy level and made it easy to change—both of which significantly increase the acceptability [11, 13]. Second, the uncertainty and severity of COVID-19 made the users extremely concerned for their and their family's health, leading to a higher (perceived) utility for sharing their data [24]. This was also evident from our study, where users opted for a lower privacy level when they felt symptomatic, possibly exposed, or were concerned enough to get a COVID-19 test. Consequently, 89% users opted for the default or lower privacy level.

The differences in privacy preferences that we observed across countries and states, we believe, were chiefly due to the varying peoples' concerns about privacy, trust in the effectiveness of the tracking technology, and the severity of the pandemic. For instance, as per a 2020 survey, 85% of Australian wanted tougher privacy laws despite having major reforms to the Australian Privacy Act in 2014 [23], four years before the European privacy law, GDPR, went into effect [12]. AU's acceptability and adoption of the technology was not as high as in the UK [26]. The pandemic in AU (in terms of COVID-19 cases and deaths) was not as severe as in the US or UK [5, 7, 19]. Thus, increased privacy concerns, medium level adoption rate [9], and relatively lesser severity of the pandemic lead to a higher privacy index for AU. On the other end, the UK was more accepting of the privacy-utility trade-off of the tracking [16], and had a relatively severe pandemic, leading to a lower privacy index. Relative to the UK, there was a lesser acceptability of the technology and much more concerns over privacy in the US [1, 24, 25, 29]. The variance among the US states can be explained in a similar way: NY and CA had a relatively severe pandemic, were more accepting of the tracking technology and less weary of privacy than FL and TX [8]. Therefore, the privacy index for NY and CA was smaller than that of FL and TX.

This study and our privacy-focused design can serve as a template for future health apps to increased users' trust and adoption. Modern privacy (e.g. differential privacy [14, 21]) assuring methods (where privacy and utility are inversely related) can exploit this design to reach an improved trade-off between privacy and the utility of the data. Furthermore, the privacy trends across age groups, gender, location as well as the predilection towards higher privacy that come with non-disclosure of information can be employed to improve the modeling user privacy profiles and automatic elicitation of privacy preferences. Hence, aiding users to protect their data and their privacy.

ACKNOWLEDGEMENTS

We thank Dr. Periklis A. Papakonstantinou (MSIS Department), Dr. Stephanie Shiau (Department of Biostatistics and Epidemiology), and Dr. Vivek Singh (Department of Library and Information Science) from Rutgers University for their help in designing the questionnaire. Research reported in this publication was supported by the National Institutes of Health (award# R35GM134927) and the National Science Foundation (award# CNS-2027789). The content is solely the responsibility of the authors and does not necessarily represent the official views of the agencies funding the research.

REFERENCES

[1] Samuel Altmann, Luke Milsom, Hannah Zillessen, Raffaele Blasone, Frederic Gerdon, Ruben Bach, Frauke Kreuter, Daniele Nosenzo, Séverine Toussaert, Johannes Abeler, et al. 2020. Acceptability of app-based contact tracing for COVID-19: Cross-country survey study. *JMIR mHealth and uHealth* 8, 8 (2020), e19857.

[2] Hafiz Asif, Periklis Papakonstantinou, Stephanie Shiau, Vivek Singh, and Jaideep Vaidya. 2022. Intelligent Pandemic Surveillance via Privacy-Preserving Crowdsensing. *IEEE Intelligent Systems* (2022).

[3] Privacy at Covid Nearby. 2022. Privacy at Covid Nearby. https://covidnearby.org/privacy-qdocs/. Accessed: 07-25-2022.

[4] Minyvonne Burke. 2020. New Jersey coronavirus patient gives hospital fake name, address — and then leaves. https://www.nbcnews.com/news/us-news/new-jersey-coronavirus-patient-gives-hospital-fake-name-address-then-n1162076.

[5] Covid-19 Confirmed Cases by Region and Time. 2022. https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports

[6] Jennifer K Carroll, Anne Moorhead, Raymond Bond, William G LeBlanc, Robert J Petrella, and Kevin Fiscella. 2017. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. *Journal of medical Internet research* 19, 4 (2017), e5604.

[7] Coronavirus cases:. 2022. https://www.worldometers.info/coronavirus/

[8] Eugene Y Chan and Najam U Saqib. 2021. Privacy concerns can explain unwillingness to download and use contact tracing apps when COVID-19 concerns are high. *Computers in Human Behavior* 119 (2021), 106718.

[9] COVIDSafe Assessment 1: National COVIDSafe Data Store Access Controls. 2021. https://www.oaic.gov.au/privacy/privacy-assessments/covidsafe-assessment-1-national-covidsafe-data-store-access-controls#_ftnref2

[10] Shai Davidai, Thomas Gilovich, and Lee D Ross. 2012. The meaning of default options for potential organ donors. *Proceedings of the National Academy of Sciences* 109, 38 (2012), 15201–15205.

[11] Isaac Dinner, Eric J Johnson, Daniel G Goldstein, and Kaiya Liu. 2011. Partitioning default effects: why people choose not to choose. *Journal of Experimental Psychology: Applied* 17, 4 (2011), 332.

[12] EU General Data Protection Regulation (GDPR). 2018. *2018 reform of EU data protection rules*. European Commission. https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf

[13] Jon M Jachimowicz, Shannon Duncan, Elke U Weber, and Eric J Johnson. 2019. When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy* 3, 2 (2019), 159–186.

[14] Zach Jorgensen, Ting Yu, and Graham Cormode. 2015. Conservative or liberal? Personalized differential privacy. In *2015 IEEE 31St international conference on data engineering*. IEEE, 1023–1034.

[15] Gabriel Kaptchuk, Daniel G Goldstein, Eszter Hargittai, Jake Hofman, and Elissa M Redmiles. 2020. How good is good enough for COVID19 apps? The influence of benefits, accuracy, and privacy on willingness to adopt. *arXiv preprint arXiv:2005.04343* (2020).

[16] Stephan Lewandowsky, Simon Dennis, Andrew Perfors, Yoshihisa Kashima, Joshua P White, Paul Garrett, Daniel R Little, and Muhsin Yesilada. 2021. Public acceptance of privacy-encroaching policies to address the COVID-19 pandemic in the United Kingdom. *Plos one* 16, 1 (2021), e0245740.

[17] Michelle Madejski, Maritza Lupe Johnson, and Steven Michael Bellovin. 2011. The failure of online social network privacy settings. (2011).

[18] Brigitte C Madrian and Dennis F Shea. 2001. The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly journal of economics* 116, 4 (2001), 1149–1187.

[19] Covid-19 Map. 2022. https://coronavirus.jhu.edu/map.html

[20] Brian W Matthews. 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure* 405, 2 (1975), 442–451.

[21] Frank McSherry and Kunal Talwar. 2007. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. IEEE, 94–103.

[22] Covid Nearby. 2020. NSF sponsored initiative by Rutgers University. https://covidnearby.org/

[23] History of the Privacy Act. 2022. https://www.oaic.gov.au/privacy/the-privacy-act/history-of-the-privacy-act

[24] Jinkyung Park, Eiman Ahmed, Hafiz Asif, Jaideep Vaidya, and Vivek Singh. 2022. Privacy Attitudes and COVID Symptom Tracking Apps: Understanding Active Boundary Management by Users. In *International Conference on Information*. Springer, 332–346.

[25] Pew Research Center. 2020. *Most Americans don't think cellphone tracking will help limit COVID-19, are divided on whether it's acceptable*. Technical Report. Washington, D.C. https://www.pewresearch.org/fact-tank/2020/04/16/most-americans-dont-think-cellphone-tracking-will-help-limit-covid-19-are-divided-on-whether-its-acceptable/.

[26] Emily Seto, Priyanka Challa, Patrick Ware, et al. 2021. Adoption of COVID-19 contact tracing apps: a balance between privacy and effectiveness. *Journal of medical Internet research* 23, 3 (2021), e25726.

[27] Haohuang Wen, Qingchuan Zhao, Zhiqiang Lin, Dong Xuan, and Ness Shroff. 2020. A study of the privacy of covid-19 contact tracing apps. In *International Conference on Security and Privacy in Communication Systems*. Springer, 297–317.

[28] Simon N Williams, Christopher J Armitage, Tova Tampe, and Kimberly Dienes. 2021. Public attitudes towards COVID-19 contact tracing apps: A UK-based focus group study. *Health Expectations* 24, 2 (2021), 377–385.

[29] Baobao Zhang, Sarah Kreps, Nina McMurry, and R Miles McCain. 2020. Americans' perceptions of privacy and surveillance in the COVID-19 pandemic. *Plos one* 15, 12 (2020), e0242652.

A APPENDIX

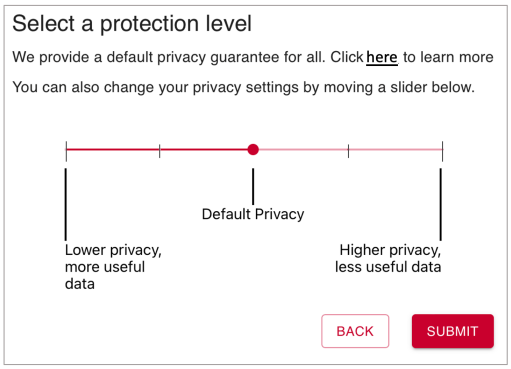

Figure 5: Privacy preference question.