

# Achieving Resilience: Data Loss and Recovery on Devices for Personal Use in Three Countries

JULIA WUNDER, IT Security Infrastructures Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany

RICK WASH, University of Wisconsin – Madison, United States

KAREN RENAUD, Department of Computer and Information Sciences, University of Strathclyde, United Kingdom

DANIELA A. OLIVEIRA, National Science Foundation, United States

ZINAIDA BENENSON, IT Security Infrastructures Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany

Recovery from adverse incidents, such as accidents or cyber attacks, is a cornerstone of cyber resilience. Backups are essential in facilitating systems recovery. We have limited understanding of how devices for personal use are backed up, and of how data loss and recovery occur, including which factors might be helpful to afford resilience. To gain insights, we surveyed almost representative (in age and gender) samples of German, UK and USA populations, 1423 in total. Almost half of the participants (656, 46%) experienced at least one data loss incident. Whereas 42% of 656 participants recovered using backups, over half of them had outdated or incomplete backups. High levels of stress were reported, especially by those recovering without backups or with problematic backups. In the full sample, 86% of participants created full or partial backups of at least one of their devices, the most important trigger being prior data loss experiences.

CCS Concepts: • **Security and privacy** → **Usability in security and privacy**; • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: resilience, backup, recovery, mobile, laptop, desktop

## ACM Reference Format:

Julia Wunder, Rick Wash, Karen Renaud, Daniela A. Oliveira, and Zinaida Benenson. 2025. Achieving Resilience: Data Loss and Recovery on Devices for Personal Use in Three Countries. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26-May 1, 2025, Yokohama, Japan. ACM, New York, NY, USA, 36 pages. <https://doi.org/10.1145/3706598.3714202>

## 1 INTRODUCTION

A resilient object is able to recover its shape and size after being deformed; it bounces back as it was. Cyber resilience, then, manifests when a person or organization is able to “bounce back” and recover the ability to function normally after a cyber attack or other adverse technical incident. Measures to facilitate cyber resilience should be a key component of any cybersecurity strategy. When *adverse incidents* occur, effective cybersecurity assures resilience and allows people and organizations to recover to their full pre-incident state. Resilience comes into play in at least two situations: (1) *cyber attacks*, including ransomware, spyware, and malware, and (2) *benign failures*, such as hard drive failures, buggy software updates, user mis-configurations and mishaps. Making and maintaining high-quality backups of both,

---

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).

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

user-generated content and functionality, is one of the most common and effective ways of ensuring full restoration of system use after adverse incidents.

In this work, we investigate the cyber resilience of computing systems utilized for personal use across three countries: the USA, the UK and Germany. Nowadays, there are many ways to back up data, for example, using the cloud or external hard disks, backing up important files manually or backing up the whole system with the help of specialized software. However, which method works best for users can depend heavily on factors such as time, money or personal experience. It is also important to examine what obstacles users face when creating a backup and when trying to recover after an incident. A backup only makes sense if the data can be restored during a recovery process. We summarize these aspects in the following exploratory research questions:

**RQ1:** Which experiences did non-expert users undergo with data loss and recovery incidents?

**RQ2:** What was the role of backups in data (content) and system (functionality) recovery?

**RQ3:** What are the current backup practices of non-expert users?

**RQ4:** Which factors influence making backup?

Although we assumed that there may be differences between countries in backup and recovery behavior, and discovered such differences during analysis, a deep cross-country analysis is out of scope of this exploratory study, as explained in more detail in Section 3.3.

We first describe end user experiences during system recovery. Many of the users we surveyed reported unusable and unsatisfactory experiences during recovery attempts. This suggests that system recovery is currently non-trivial. We further describe patterns in what users choose to back up, and therefore, which parts of their systems are definitely resilient<sup>1</sup>. We find generally high levels of backup use (86% of respondents backed up *something*), though once we dig deeper, we find that many did not back up all of their devices (i.e., they did not backup *anything* from these devices), nor all of their data. While there may be different reasons for this behavior, one of the key challenges in fostering cyber resilience remains ensuring that usable backups are made that can enable recovery in the face of adverse incidents. The main contributions of our study are to describe how recovery accidents happen, and how backups are being done in our sample in the three countries. Our main findings are as follows:

- Recovery experiences:
  - Needing to recover from data loss is common, with 46% of participants reporting at least one incident.
  - Recovery was often fraught and stressful (over 70% said so), but it was especially stressful in the absence of backups, or if existing backups turned out to be problematic (outdated, incomplete, not working).
  - In many cases, recovery was incomplete and resilience suboptimal, as 70% of the participants who attempted to recover only achieved partial recovery of lost data.
- Making backups:
  - A majority of participants (86%) made some kind of backups of at least one of their devices, suggesting that the need to do this has been effectively communicated to the general population.
  - People understood the need to back up their data, but there were impediments to actually creating backups such as lack of knowledge, perceived unnecessary, financial constraints, and time limitations. Especially, 15-40% of participants (depending on device) who did not back up their devices, agreed that backups are necessary for their respective devices.
  - Participants mostly start saving their data as a backup when they have experienced a data loss.

<sup>1</sup>We recognize that there are also other ways to maintain resilience, as we explain in Section 2.1.

- Backup engagement was not evenly distributed among users: Older users were more likely to back up laptops and desktops, whereas younger users were more likely to back up their smartphones.
- For desktops and laptops, there is almost no difference between genders; both men and women are equally likely to back up their devices. However, when it comes to phones, women are much more likely to back up their phones than men.
- Individuals with higher incomes are more likely to back up their laptops and smartphones.
- Desktop computers are more likely to be backed up by persons that have other people in the house besides their partner. For phones, individuals are more likely to backup if they have other people in the house, inclusive of their partner.
- Backups frequently included user generated content (content backup), but less frequently included data necessary for restoring system functionality (functionality backup) such as executables, configuration, etc.

We note that our findings refer only to the three studied countries, and may not be transferable to other countries due to differences in cultures, demographics or infrastructure.

The rest of this paper is structured as follows: we commence by discussing key terms and reviewing related research in Section 2. We then detail the study design in Section 3 and report on findings in Sections 4 and 5. Section 6 discusses our findings and Section 7 concludes.

## 2 BACKGROUND

### 2.1 Delineating Key Terms

*Resilience* can be defined as “the ability of a system or network to withstand and recover from disruptions” [35, p.274]. *Recovery* in the cyber realm includes recovering lost data and functionality after an incident, which relies heavily on backups. *Backups* are an independent storage of data copies, secured with the aim of being able to restore lost data to support recovery from adverse incidents [5].

Resilience as defined above is not concerned with confidentiality. Usually, the loss of confidentiality cannot be recovered, as the information is already in the wrong hands. Whitten and Tygar call this property of confidentiality the “barn door” property of security [38].

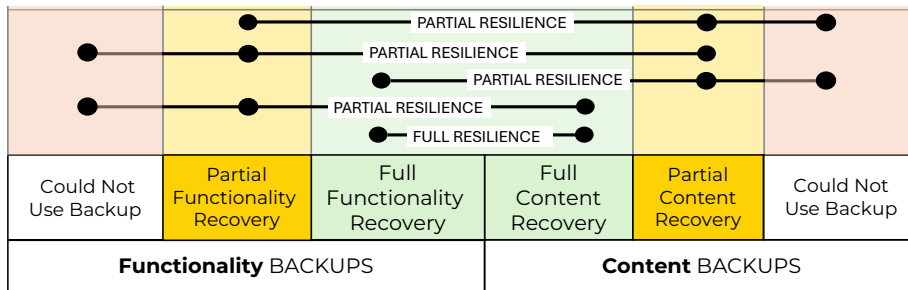


Fig. 1. Conceptual Model of the Role of Backups in Facilitating Resilience.

As shown in Figure 1, resilience traditionally depends on being able to recover from adverse incidents, and recovery relies on the data owner having made backups of content and functionality before the incident<sup>2</sup>. True resilience is

<sup>2</sup>This figure does not reflect the situation when a hardware failure has occurred which might require a purchase of new hardware before recovery can commence. Here, we only reflect on the role of backups in affording resilience.

elusive if the recovery process is incomplete or infeasible. The diagram highlights the essential role of backups in the resilience context. Although full recovery might also be possible without backups, its success in this case heavily depends on what happened. For example, if the system was attacked by ransomware, it is possible to get files back by paying, or by exploiting a mistake in encryption that the criminals made. If the boot sector or other small part of the hard disk is affected, it might be possible to get the data back using special equipment, and reinstall the needed software. In contrast, with backups that are complete and could be used, resilience is guaranteed.

There are economic trade-offs in resilience. As the backup of the full system (content and functionality) incurs non-negligible costs, users may decide not to do the full backup, especially if some parts of their system are not important to them, or can be recovered through purchases or downloads of new copies of data and software. Thus, a decision not to backup unimportant pictures, or operating system and movies that can be re-purchased, is not bad per se, as long as it can be economically justified, and especially the costs of losing some data forever and recovering without backup are factored in.

## 2.2 Related Work

Creating and maintaining backups is a key cybersecurity measure which enables an effective adverse incident response. Our particular interest is in devices for personal use and users' resilience and, in particular, their recovery from adverse incidents. This may include using backups if these exist. The US company Backblaze [39] reported in 2022 that only 10% of their respondents backed up daily, with 35% never making backups. Cooke [8] reported in 2024 that there is a lack of education on backup technology.

Wash and Rader [36] asked a representative sample of the US population about their security behaviors. Only 24.2% of participants reported that they regularly engaged in advanced security behavior, such as applying patches, disabling scripts in web browsers and making backups<sup>3</sup>. It is possible that people do not know that they should make backups, i.e., they are not being made aware, it could mean that they are aware of the need for backups but do not have the capacity or skills. Even if they know they should, and know how, they might not be able to afford a separate storage mechanism. Awareness of backups is clearly the first and essential pre-requisite for backups to be made, but many additional factors may play a role.

A number of research publications provide advice specifically to organizations [2, 12, 15], but research on specific backup advice targeting devices for personal use is sparse. Users might well look to governments for cybersecurity advice, and governments do indeed include backup advice in their targeted advice (e.g., German Federal Office of Information Security [10], the UK's National Cybersecurity Centre [18] and the USA's CISA [30]). However, there is some evidence that users might not consult their government's publications for cybersecurity advice [21]. Low national trust levels<sup>4</sup> may prevent citizens from consulting and/or following advice provided by their government [17].

Users also likely rely on search engines or media to get cybersecurity advice in general [20, 26]. However, there is a distinct lack of agreement between experts on cybersecurity measures to be taken, and especially on their prioritization [4, 11, 27–29]. This disagreement might lead to information overload, uncertainty and perhaps result in inaction. Moreover, only two of the above publications mention backups in the context of provided advice: Redmiles et al. [28] point to backup-related advice being in need of improvement, and Reeder et al. [29] include admonitions to backup that 10 out of 231 of their expert respondents mentioned. Considering the current state of play, it seems likely that few

<sup>3</sup>No separate statistics for backup are presented in [36]

<sup>4</sup>The Edelman Trust Barometer [9] reports the following citizen government trust levels for 2024: UK: 39%; USA: 45%; Germany: 46%

individuals actually get the right advice that will lead them to contemplate their own resilience, anticipate having to recover and consequently make backups to facilitate recovery should the need arise.

Research into how end users make backups appears sparse, with some notable exceptions. Kljun et al. [13] found that the majority of their participants made manual, selective, and noncontinuous backups of their computers. They also reported that a fifth of their respondents' computers were not backed up at all, meaning that recovery and resilience are rendered challenging. Some individuals might pay a service or private person for help after an adverse incident. This might put them at risk of privacy violation and potential data theft. Although we are not aware of research into such situations in Germany, the UK and the USA, this seems to happen in Bangladesh according to Ahmed et al. [1]. Although cultural, religious and legal contexts are quite different in these countries, the threat to privacy cannot be ruled out.

Muslukhov et al. [16] interviewed 22 participants at a university in Canada to explore users' requirements for securing data on their smartphones, including making backups. Half of the participants relied on external hard drives to backup their personal files, which include videos, pictures and documents. Although many of them also used cloud storage services, they preferred to save only shareable content on the cloud, as most participants mistrusted the cloud providers when it came to handling sensitive data. Additionally, high costs, limited availability and slow access speeds are also seen as challenges. Menard et al. [14] investigated the intention of adopting cloud backup and the factors influencing this decision by using theoretical scenarios. They conducted an online survey with 152 participants from a US university, presenting various scenarios to measure their willingness to backup data in the cloud. The findings revealed that perception of threats as more serious and likely, as well as perception of cloud as convenient led to higher intention to use cloud for backups. Ultimately, convenience had a greater impact than perceived threats, and demographic data did not have influence. Redmiles and Hargittai [25] investigated US students' engagement with backing up their computers and smartphones. They found that less than half of their participants made regular backups and also uncovered a strong relationship between Internet skills and backup frequency.

Given that this research was published some years ago, our study could provide novel insights in a fast-changing domain. We mainly drew on the related work of Redmiles and Hargittai [25] and Kljun et al. [13] to define some questions about backup behavior and backup methods. However, in contrast to these papers, we focus on four device types for personal use (desktops, laptops, smartphones and tablets) and use representative in age and gender samples from three countries. Kljun et al. [13] considered non-mobile devices in a convenience international sample, and Redmiles and Hargittai [25] considered computers and smartphones in a US student panel. Moreover, almost all mentioned papers focus on backups and do not delve into recovery experiences, apart from Kljun et al. [13] who present a very short section on "backup stories". This section asserts that backing up and recovering was often difficult due to not understanding how technology works, technology failure or user mistakes. We think that a deeper understanding of backup and recovery behavior would deliver valuable insights not only for those formulating advice for end users, but also for cloud providers, device and operating system manufacturers and recovery software providers.

### 3 STUDY DESIGN

#### 3.1 Questionnaire

Figure 2 shows a simplified version of the questionnaire. The full questionnaire can be found in Supplementary Materials. The questionnaire consisted of two main sections: (1) Backups and (2) Data Loss/Recovery. As mentioned in the previous section, we used related work by Kljun et al. [13] and Redmiles and Hargittai [25] to get inspiration for questions to ask

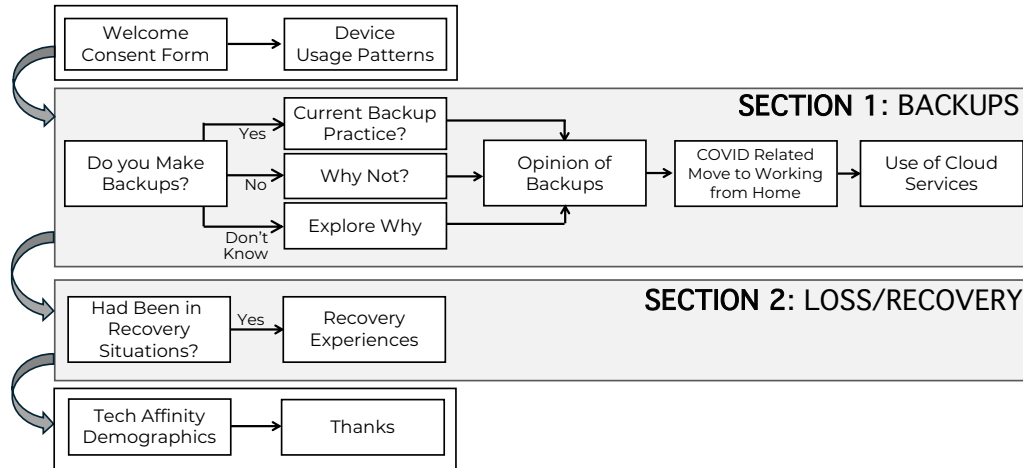


Fig. 2. Schematic representation of the questionnaire; full questionnaire can be found in Supplementary Materials.

and factors to investigate about backup. Furthermore, we used related work by Rader et al. [24] about security stories described by end users in a survey, and the stories from the paper published in OSF (Open Science Framework) [23] to understand how best to ask about recovery stories, and which wording to use and to expect from participants.

After providing informed consent, the participants were asked which devices (desktop computer, laptop, smartphone, tablet) they use, which operating systems (OS) they use, whether they use the devices for personal or work purposes and how long they have been using the devices. To keep the questionnaire as short as possible, we asked the participants to indicate only one non-mobile device (laptop or desktop). Considering mobile devices (smartphone and tablet), we asked the participants to indicate both devices if applicable. If they had more than one device of any type, we asked them to select the one they utilize most for personal use.

*(1) Backup Section.* This section considered current backup strategies separately for each device. If the participants did not make backups, they were asked for the reasons, such as, they did not know how to make backups, or considered backups unnecessary for this particular device. If participants made backups, they were queried about the exact backup methods on three dimensions: (1) where the backup resided (e.g., cloud, external hard drive, USB drive, email attachment, etc.), (2) which data they saved, in detail, and (3) how often they made the backup. Then, allowing a free text response, we offered participants the opportunity to comment further on their backup preferences if they so desired. Next, we asked why they started backing up and whether they regularly checked their backups to make sure they were working properly. Then we offered the participants various statements to determine agreement, such as “I have backed up all important and relevant files” or “My backups are up-to-date”, using a 5-point Likert scale. We also asked whether the participants had ever experienced problems with backups, such as an incomplete backup.

The survey then elicited participants’ opinions on backups, independently of whether they make backups or not. Here, we included the first of two attention tests. We asked the participants to provide a free text response explaining the characteristics of a backup method they considered important. This was followed by questions related to the COVID-19 pandemic: whether the participants had worked from home during the pandemic, whether they had used personal or business devices and whether their backup behaviors had changed as a consequence of home working during the

pandemic. The next part dealt with the use of cloud services: whether they used these services, and, if so, which cloud providers, for what purpose they stored files and whether they thought the cloud services created backups of their data. Finally, all participants were asked whether they would like to back up all their data, applications and settings to the cloud and were requested to justify their response in a free text field.

(2) *Data Loss and Recovery Section.* Next we posed questions about data loss and recovery experiences. Participants were first asked whether they had ever experienced data loss. If not, subsequent questions were skipped. If they had, they were asked how often such an incident had occurred. Then, participants were asked to briefly describe the incident they remembered best. This was followed by further questions about the incident, such as how long ago it occurred, which device was affected, and what the possible cause might have been. We then asked if the participants had contacted others for assistance. If not, they were asked why. If they did seek help, they were asked whom they consulted, whether they had to pay for the service, and, if so, how reasonable they felt the cost was. We also asked whether they had had privacy concerns, in terms of giving helpers access to their devices.

We then posed questions about the measures they took after the incident, such as recovering data from backups. If a backup had been used, we asked if there were any problems during the recovery process. We then queried whether any data had been irretrievably lost due to the incident and, if so, what the nature of data was (such as personal photos or documents).

We did not ask what data was initially affected quantitatively, but instead asked what could not be recovered (see Section 4.4). We did this to avoid ambiguity, as lost and later recovered data does not lead to the same outcome as data lost forever. We were especially interested in impediments to the recovery process, and in data that could not be recovered at all, meaning a loss of resilience.

This was followed by a question asking how long it took before they could use the affected device again as usual, if at all. Next, we posed statements about the recovery process, such as “It was emotionally distressing” or “It took a lot of effort”, to which they could indicate agreement on a 5-point Likert scale. The second attention test was included here. Finally, we asked the participants whether their backup behaviors had changed as a result of the incident, e.g., whether they started or ceased making backups.

The survey concluded with questions about participants’ technical affinity and demographics.

*Translation and Testing.* The questionnaire was finalized in English. Because German participants were also being recruited, the questionnaire was subsequently translated into German. Three researchers independently translated the questionnaire, which was then discussed iteratively in several sessions until a unified version was agreed upon. Next, the language service center of our university provided a translation from German to English, which was compared to the original questionnaire for semantic equivalence. The only question that differs from country to country is income. We used the Atlas method [3] to convert the scale from US dollars to British pounds for the UK and Euros for Germany. The scale for Germany was then further adapted to inquire about monthly instead of annual income, as this is more common in Germany.

Before the questionnaire was published, there were several rounds of testing, with the questionnaire being iteratively adapted. Four research colleagues working in the field of IT security took part in the first round of testing. This was followed by a second round of testing with 7 participants from the circle of acquaintances, where care was taken to recruit testers who did not work in IT or science. This second round of testing took place as a think-aloud test via Zoom using screen sharing so that all feedback could be recorded. In a third round of testing, we recruited around 90 participants via Prolific (28 USA, 30 UK) and Clickworker (30 Germany) to test the functionality of the questionnaire.

Because the questionnaire did not change after the third round of testing, this was used as the final version. Participants of the first two test rounds received no compensation, participants of the third round were compensated \$3 on Prolific and 3€ on Clickworker.

### 3.2 Data Analysis

*Qualitative analysis:* Within the survey, there were four large free text fields that were evaluated qualitatively: 1) a list of backup characteristics that are important to the participants; 2) an explanation of whether the participants wanted to upload all their data to the cloud, or not; 3) descriptions of incidents during which participants lost data and wanted to recover it; 4) anything else the participants wanted to share about their backup practices. For each of these four fields, two researchers independently created codebooks. The first 120 free text answers (40 per country) were then coded. In several sessions, the coding of these comments was discussed by the two researchers and the codebook was adjusted until a final version was created that left no room for ambiguities. One of the researchers then coded the remaining comments, a total of around 600-1400 comments per free text field. The final coding was discussed and interpreted during team meetings.

*Quantitative analysis:* The majority of the questions were multiple choice. We report tallies and percentages of respondents for these questions. Many of our findings are based on counting responses to questions.

Moreover, we conducted a series of logistic regression models to look at the relationship between backing up data (the dependent variable) with a number of variables indicating prior experiences and/or demographics. We report logistic regression coefficients and whether each one is statistically significant at the 5% level. A small number of individuals who did not answer any of the predictor questions used in the regression were removed for these regressions (due to missing data). For categorical variables with no natural baseline (like *Gender*), the omitted level is the baseline (*Female*).

To interpret these regression coefficients we often generated “predicted values” by using the regression model to estimate the likelihood that a statistically average person with specific characteristics would back up their data; these predicted values make it easier to understand the magnitude of differences we observed. These predicted probabilities are not actual individuals, but rather estimated statistical averages.

### 3.3 Ethics and Recruitment

The data protection officer of Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany, approved the study design, as the university does not provide an ethical committee for non-medical studies. We assumed that backup and recovery behavior may be different by country due to the differences in cultures, demographics, infrastructure, digitization grades and other factors. Therefore, we decided to recruit participants from three countries: Germany, the UK and the USA. However, we did not know which differences we are going to find, and therefore, did not aim at a cross-country analysis in this exploratory study, as reflected in our RQs (Section 1). The reason is that without knowing concrete differences, it is difficult to predict which factors may be behind them. Thus, we decided to focus on gathering facts first, leaving a deeper cross-country analysis to future work.

We further assumed that backup and recovery behavior may be different by demographics, although we did not know how exactly. Therefore, we decided to recruit samples that are representative in age and gender. In this exploratory study, we decided not to consider representativeness by other factors, such as education or income, because this would place too many restrictions on the target recruiting platforms.

Table 1. Demographic distribution of gender and age in Germany [19], number of associated recruitment slots and participants actually recruited. We aimed to recruit 500 participants.

	Distribution in Germany in %		Number of recruitment slots		Number of recruited participants	
	Male	Female	Male	Female	Male	Female
18-27 years	6.8%	6.2%	33	31	33	31
28-37 years	8.0%	7.5%	40	38	40	38
38-47 years	7.3%	7.3%	38	36	38	36
48-57 years	9.0%	8.9%	45	44	45	44
58 years or older	17.9%	21.2%	89	106	72	48

We aimed to recruit 500 participants per country. The UK and the USA participants were recruited via Prolific, where we recruited an almost representative sample, based on the distribution of gender, age and ethnicity in the local Census, using a special option offered by Prolific. As the German worker community was deemed inadequate on Prolific, the Clickworker platform was used for Germany<sup>5</sup>. Clickworker did not support selection of a representative sample, so we created several participant groups in order to recruit a sample that is was representative as possible based on the distribution of gender and age in Germany (see Table 1).

Participants could terminate the survey at any time or contact us via email. At the beginning of the survey, participants were informed about their data protection rights. Only participants who provided informed consent could take part in the survey. The participants were compensated according to the recommendation of the platforms and received a payment of \$3 on Prolific and 3€ on Clickworker for the 15-minute survey. Only fully completed surveys were compensated. The collected data was stored on a secure server in the first author’s lab, where only employees and project members had access. The survey was conducted on a self-hosted LimeSurvey<sup>6</sup> instance at IT Security Infrastructures Lab of FAU.

### 3.4 Participants

Although we aimed to recruit around 500 participants per country, it proved difficult for Germany. We made several calls on the Clickworker platform, but were unable to find enough older participants, such that there are fewer German participants than expected. For example, we aimed to find 106 female German participants over 57 in order to gain an almost representative sample in age and gender (see Table 1, 21.2% of 500 is 106). However, after several calls, we recruited 48 female German participants over 57. A few participants were excluded because they did not pass the attention tests. A total of 1423 participants provided eligible answers: 494 from the USA, 498 from the UK and 431 from Germany.

An overview of the participants’ demographics is presented in Table 2. Approximately half of the participants self-identified as male and the other half as female. On average, participants were 46 years old at the time of the survey (median: 45), with the youngest participant being 18 and the oldest 86 years of age. About 43% of the participants did not have an academic degree, about a third had a Bachelor’s degree, 25% a Master’s degree or higher. Half of the participants were employed, 18% self-employed and 10% retired. The majority of participants (64%) had a long-term partner. About half of the participants had children, with about 24% living with their children, and 21% percent of

<sup>5</sup>At the time of the survey, about 3,000 German participants were available on Prolific who had been active in the last 90 days [22] and about 540,000 German participants were registered on Clickworker [7].

<sup>6</sup><https://www.limesurvey.org>

Table 2. Overview of the participants' demographics; the percentages within the columns refer to the percentages of the respective country, "CS" means computer science.

	All		USA		UK		Germany	
	N	%	N	%	N	%	N	%
Total	1423	100	494	100	498	100	431	100
Male	697	49.0	230	46.6	238	47.8	229	53.1
Female	694	48.8	252	51.0	246	49.4	196	45.5
Diverse	13	0.9	6	1.2	5	1.0	2	0.5
N/A	19	1.3	6	1.2	9	1.8	4	0.9
18-29 years	278	19.5	102	20.6	104	20.9	72	16.7
30-49 years	492	34.6	170	34.4	167	33.5	155	36.0
50-59 years	289	20.3	92	18.6	88	17.7	109	25.3
60 years or above	343	24.1	126	25.5	128	25.7	89	20.6
N/A	21	1.5	4	0.8	11	2.2	6	1.4
No academic education	611	42.9	210	42.5	196	39.4	205	47.6
Bachelor's degree	432	30.4	189	38.3	187	37.6	56	13.0
Master's degree	287	20.2	75	15.2	77	15.5	135	31.3
Ph.D.	62	4.4	17	3.4	26	5.2	19	4.4
Other	13	0.9	0	0.0	4	0.8	9	2.1
N/A	19	1.3	3	0.6	8	1.6	8	1.9
Pupil, student, apprentice	87	6.1	15	3.0	34	6.8	38	8.8
Employee, civil servant	736	51.7	268	54.3	256	51.4	212	49.2
Self-employed, freelancer	252	17.7	81	16.4	78	15.7	93	21.6
Unemployed	58	4.1	35	7.1	17	3.4	6	1.4
Homemaker	63	4.4	19	3.8	25	5.0	19	4.4
Pensioner	149	10.5	37	7.5	66	13.3	46	10.7
Other	48	3.4	26	5.3	14	2.8	8	1.9
N/A	18	1.3	1	0.2	8	1.6	9	2.1
Partner	899	63.7	291	58.9	349	70.1	259	60.1
Children	683	48.0	236	47.8	264	53.0	183	42.5
Children in household	337	23.7	127	25.7	125	25.1	85	19.7
Living alone	297	20.9	108	21.9	79	15.9	110	25.5
CS-related work	228	16.0	97	19.6	59	11.8	72	16.7
Non CS-related work	1195	84.0	397	80.4	439	88.2	359	83.3

participants lived alone. In terms of computer skills, most considered themselves fairly (55%) or very (23%) skilled, with 16% working or studying in the field of computer science or a related field.

#### 4 DATA LOSS AND RECOVERY

Almost half of all participants (46%, 656 out of 1423) had experienced data loss and wanted to recover data. Of these, around half (49%) stated that such an incident had occurred once. The other half lost data two (28%) or more (23%) times. This constitutes a substantial number of data loss incidents, and shows that cyber resilience is more than a hypothetical concern; it is a real part of many people's lives.

We asked the 656 participants, who experienced data loss at least once, to describe a single specific incident where data was lost that they could remember best, which was followed by quantitative questions about the incident. We now

Table 3. Overview of when the incident occurred and which device was affected. The percentages within the columns refer to the percentages of the affected devices.

	<b>Desktop Computer</b> <i>N</i> = 253	<b>Laptop</b> <i>N</i> = 231	<b>Smartphone</b> <i>N</i> = 157	<b>Tablet</b> <i>N</i> = 9
Within last month	2.0%	4.3%	8.3%	11.1%
Within last year	12.3%	15.6%	14.6%	33.3%
Within last 3 years	20.9%	26.0%	24.2%	33.3%
More than 3 years ago	64.8%	54.1%	52.9%	22.2%

highlight these incidents and the recovery stories they told. An overview of the codes for 656 free text responses is provided in Table 11 and Table 12 in Appendix A.

#### 4.1 What Happened?

Most participants described an incident that occurred more than three years before (57%), though many chose to describe one that occurred within the last three years (23%) or within the last year (19%), see Table 3. Most affected devices were desktop computer (39%) or laptop (35%). *“My computer broke and I lost all of my notes, including the work that I was working on that I then had to restart, P205 referred to their recovery story. Further, 24% of the participants described incidents on smartphones that led to data loss, as P79 mentioned: “My iPhone completely bricked and I had to get a new one.”* Data loss on a tablet was only described by 1% of the participants. These responses indicate that memorable data loss incidents happen across all types of devices, though loss definitely seems more rare on tablets than other devices. To reduce complexity and avoid ambiguity, we asked the participants to describe an experience that they remember the best, and on only one device even if several were affected.

The incident was almost always originally noticed by the participants themselves (97%), though some were informed by family or friends (2%). Table 4 shows an overview about the issues that made the participants aware of the incident. About 79% of the participants experienced technical issues with their device, with 39% specifically attributing it to a broken device. Lost device functionality, additionally to the missing data, is often a feature of these data loss incidents.

Only 6% stated that it was due to a virus, as P25 mentioned: *“My desktop got a virus a long time ago and I lost everything.”* Furthermore, 18% participants did not know what exactly happened to their device, which was summarized by P381: *“I’m really not sure what happened with my very first laptop, but it died and I was not able to recover anything.”* While end users are mostly able to recognize that data has been lost, they sometimes struggle to identify the root cause.

However, some causes are more evident. Some participants reported that their devices were dropped, lost or stolen (6.1%, 2.7% and 3.2%, respectively). For example, P862 stated: *“My laptop was stolen with no backup done.”* Furthermore, 3% reported that they spilled liquid on their device, as P130 said: *“I spilled water on my laptop and was unable to access the hard drive.”* Additionally, 13.1% of respondents reported that they had accidentally deleted some or all files, as P44 stated: *“I deleted a photo I definitely wanted to keep,”* and P3 narrated: *“I work as a photographer, after an event one day I mistakenly formatted my memory cards without copying the pictures to my laptop.”* Finally, 8.2% indicated that something else happened, such as that the device was damaged by another person or data suddenly disappeared.

#### 4.2 Which Data Was Affected

In the free text stories, 46% explicitly mentioned that the incident affected all data on their device, as P904 reported: *“I lost my whole computer’s files once when it broke.”* *“My phone just went black and never came back on,”* P188 stated. This

Table 4. Overview of what issues made the participants aware of the incident ( $N = 656$ ). Several issues could be selected.

<b>What issues made you aware of the incident?</b>	<b>N</b>	<b>%</b>
	656	100%
<i>Technical Issue</i>		
Could not access folders or files on the device	286	43.6%
Device did not start	248	37.8%
Device behaved strangely	75	11.4%
Device slowed down	37	5.6%
Popup blocked access to files and asked for payment, update or phone call	9	1.4%
Antivirus software displayed an alert	8	1.2%
<i>Other Issue</i>		
Files were deleted accidentally	86	13.1%
Device fell down	40	6.1%
Device was stolen	21	3.2%
Liquid was spilled onto device	20	3.0%
Device got lost	18	2.7%
Something else	54	8.2%

suggests that loss of functionality is often a part of data loss, and that restoring functionality is frequently going to be as important as restoring content.

Other mentioned lost data included photos (18%), documents (11%) or other files (4.3%) such as game files or programs. *“I had an issue with a phone that caused all my files and photos to be deleted. I had about 8 months’ worth of photos on there, pictures of things that were priceless to me,”* P108 stated.

### 4.3 The Recovery Process

Recovering from data loss is often a social experience. Whereas 47% of the participants (306 of 656) consulted others for help, 49% chose not to do so, and the rest could not remember. When our respondents did not seek help, they indicated as the reasons being able to solve the problem themselves (34%) or not wanting to bother anyone (9%).

For the 47% of respondents who did consult others, that help could come from multiple sources. They consulted professional IT support (56%, 170 of 306) and/or friends and family (47%, 143 of 306). P138 summarized their recovery story as: *“We had our desktop completely crash and be taken over by a virus. I had to hire a professional, to recover all the photos and data. We had a lot of files and family pictures on there etc. Thankfully he was able to recover everything. This happened several times over many years.”*

Only 23% (71 of 306) of the participants who sought help paid for it, with most of those (68%, 48 out of 71) considering the cost to be reasonable, 27% (19 out of 71) thought the cost overly expensive and 6% (4 out of 71) thought it was a bargain. *“I considered paying for professional recovery services but could not justify the cost,”* P906 stated. Because we wanted to explore privacy concerns related to seeking help, we also asked those who sought support whether they were concerned about the person helping them accessing their personal data. However, only 19% agreed or strongly agreed with this statement. Moreover, 2% of those who did not seek help stated that privacy concerns prevented their seeking assistance. This suggests that, on the whole, recovery service and personal contacts were considered trustworthy.

Regardless of whether outside help was consulted, various methods were attempted to restore files after the incident, according to the quantitative multiple choice questions. The most common was restoring data from a backup (42%). *“I accidentally changed data in a spreadsheet and recovered this data from a recent backup of that file,”* P902 mentioned.

Almost as many respondents, though, reported buying a new device (39%) as a result of this incident. Furthermore, 34% rebooted the device, 25% reinstalled the operating system, 21% reinstalled some software and 12% ran antivirus software. *“I used a free software to get the files back, the best it could i probably saved 85% of the data,”* P788 stated. Much of the work they reported involved attempts to restore functionality (buying new device, rebooting, reinstalling software, antivirus). This suggests that while participants might have been able to recover lost content from a backup, they might not have been able to recover lost functionality.

Only 47% of the participants who restored their data from a backup did not encounter any problems during this process (131 out of 276). This is a very low number, and suggests that the usability and success of recovery from backups still needs to be significantly improved.

Various problems with backup occurred. Thus, 24% (67 out of 276) stated that the backup was outdated or incomplete, 18% (51 out of 276) that the recovery process was complicated, 16% (44 out of 276) that the process failed and 8% (23 out of 276) could no longer access the backup. P644 summarized their recovery story as: *“I had files on a laptop which began to work inconsistently. I backed up sentimental photos onto a hard drive. That hard drive then wouldn’t open and I couldn’t access my photos. The laptop stopped working completely and I lost all the photos.”* Furthermore, 6% of the participants (16 out of 276) had a backup but did not know how to recover from it. P297 stated: *“My computer started dying and I had backed up, but when the computer died I had no idea how to recover the operating system and files from the backup”*.

Some participants explicitly mentioned that recovery was infeasible because they did not make backups (10%). *“I didn’t back up my childrens baby photos and lost them all,”* P641 stated. P4 said: *“I lost my clash of clans game after 8 years of playing it because I didn’t understand how to backup my phone.”*

#### 4.4 The Aftermath

As a consequence of the incident, 70% of the participants lost data that could not be recovered. Personal photos or videos (70%), work or school documents (35%), movies or music (26%) or apps (26%) were irretrievably lost. Whereas 31% of the participants indicated that they were able to use the device again immediately after the incident, for 45% it took a few days, for 8% a few weeks, and 16% of the participants never used their device again: *“I had a laptop and it went to a black screen of death and could never be restored”*, P337 reported. This represents a large number of people who lost data permanently, despite having backups and attempting recovery.

We asked the participants to indicate the extent to which they agreed with various recovery-related statements (see Figure 3). Many agreed that the recovery process had been stressful (72%) or emotionally distressing (71%), as well as requiring a great deal of time (61%) and effort (57%). Around 5% of the free text responses highlighted emotional stress, while 5% underlined having lost particularly important data. *“I lost many photos and was very upset that I could not find a way to get them back. I never did,”* P5 stated and P161 narrated: *“My main work Mac died and I lost EVERYTHING. It took me days and days to fix it. I lost money because I couldn’t work. Never again!”*

Most participants disagreed with the statement that a lot of money was lost (74%). Similarly, most participants disagreed with the statement that the incident had a negative impact on their reputation or social relationships (80%). While recovery was stressful, it did not seem to have a serious impact on other aspects of their lives.

Figure 4 shows a comparison of the indicated stress levels of the participants to the results of the recovery process. We can see that participants whose data could all be restored felt less stress than those who lost some data forever. If

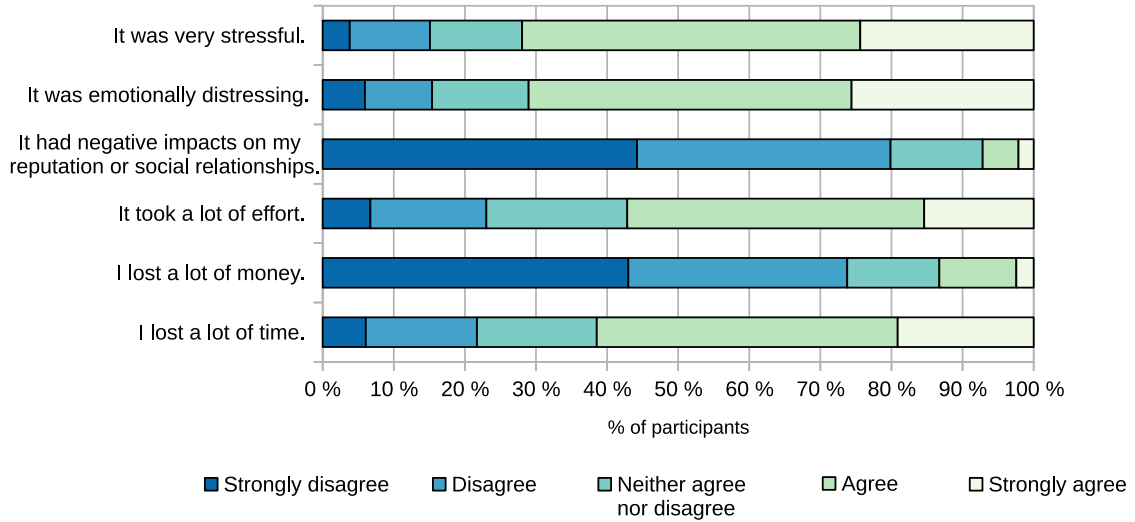


Fig. 3. Agreement to statements concerning the perception of recovery process (N = 656).

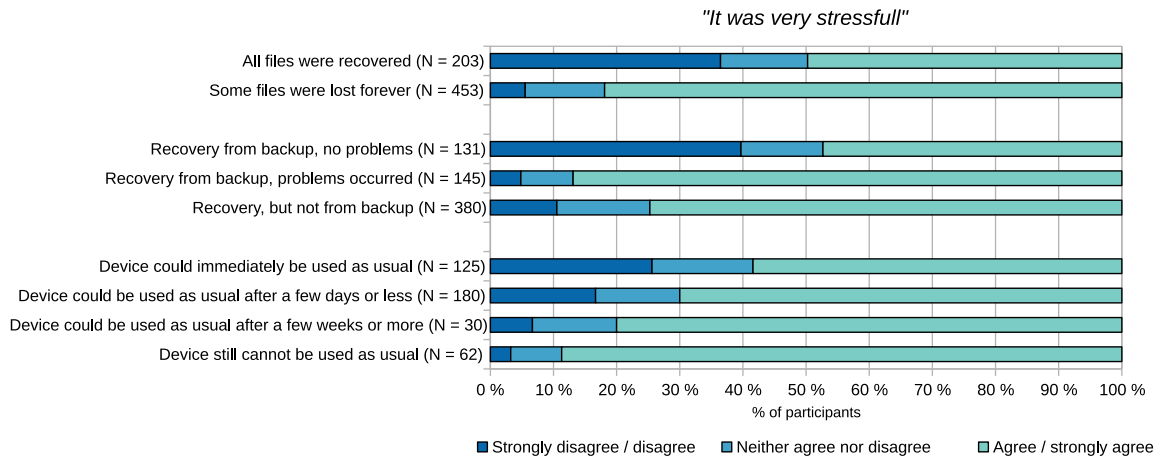


Fig. 4. Agreement with the statement "It was very stressful" compared to the indication whether the data was restored from a backup, whether all data could be restored or not and how long it took until the device could be used again as usual (N = 656).

the participants used a working backup, their recovery process was less stressful than if they did not use a backup for recovery. However, most stress was felt by people who made backup, but encountered problems with it during recovery.

Participants who were able to use the device as usual again immediately seem to have experienced less stress than those for whom it took some time. Participants were most likely to agree with the statement that the recovery was stressful if their device could not be used as usual at all. Very similar effects can be observed for the statement "It was emotionally distressing", see Figure 12 in Appendix B.

Table 5. Overview about changes in backup behavior as a result of the incident ( $N = 656$ ).

	<i>N</i>	%
<b>Did you change your backup behavior as a result of the incident?</b>	656	100%
Yes, I started doing backups	300	45.7%
No, doing backups just as before	234	35.7%
No, still not doing backups	46	7.0%
Yes, I stopped doing backups	3	0.5%
Yes, other changes, please specify	73	11.1%

#### 4.5 Changes in Backup Behavior

Around 46% of the participants started creating backups as a result of the incident (Table 5). For 43%, their backup behaviors did not change, with 36% making backups exactly as they did before and 7% continuing not to make any backups. P373 summarized their story as follows: “A long time ago, my computer crashed. Since I didn’t back it up, I lost everything. Ever since then, I have always used some type of backup method.”

Some participants changed their backup methods. The fullest account comes from P145, who switched from automatic backup of the whole system (which did not work when needed) to partial manual backups: “I think it was a windows laptop that failed catastrophically somehow. Like one day it just wouldn’t turn on or something. [...] I had been using some sort of Windows backup and the backups could not be read or could not be restored for some reason. [...] I switched to backing up the files I cared about rather than system images and stopped using that backup software.” We do not know which particular backup software this participant used, but it is clear that it did not fulfill its promise, and thus the participant abandoned it, and got disappointed in the “full backup” strategy.

## 5 CURRENT BACKUP PRACTICES

As we could see in Section 4, people generally seem to have an easier time recovering when they had made *unproblematic* backups, and seem to have the most stressful recovery if they made backups, but encountered various problems with them. Moreover, although a recovery without backup is indeed possible, it is always fraught with uncertainty, depending of which kind of problem caused the loss. In this part of the paper, we consider the current backup practices of users in Germany, the UK and the USA. We utilize almost representative population samples (in age and gender) from each country: 431, 498 and 494, respectively, 1423 in total (see Section 3.4).

Approximately 34% of participants use a desktop computer for personal use and 62% use a laptop. We asked participants to choose between one of the two devices. A smartphone is used by 93% of all participants, and 31% use a tablet for personal use. Most participants (60% - 80%, depending on device) have been using these devices for more than 3 years.

Figure 5 shows the percentage of participants who create backups of their devices. Around 86% of the participants create backups of at least one of their devices. Desktop computers (83%) and laptops (74%) are more likely to be backed up than smartphones (69%) and tablets (53%). The differences between the three countries are small, mostly around 5 pp. The most notable difference is in smartphone backup: Whereas in the USA, 74% of participants backup their smartphone, in the UK it is 68%, but in Germany it is only 54%. Overall, the three countries are quite similar, such that in this exploratory study, we mostly analyze the data together to have a large sample.

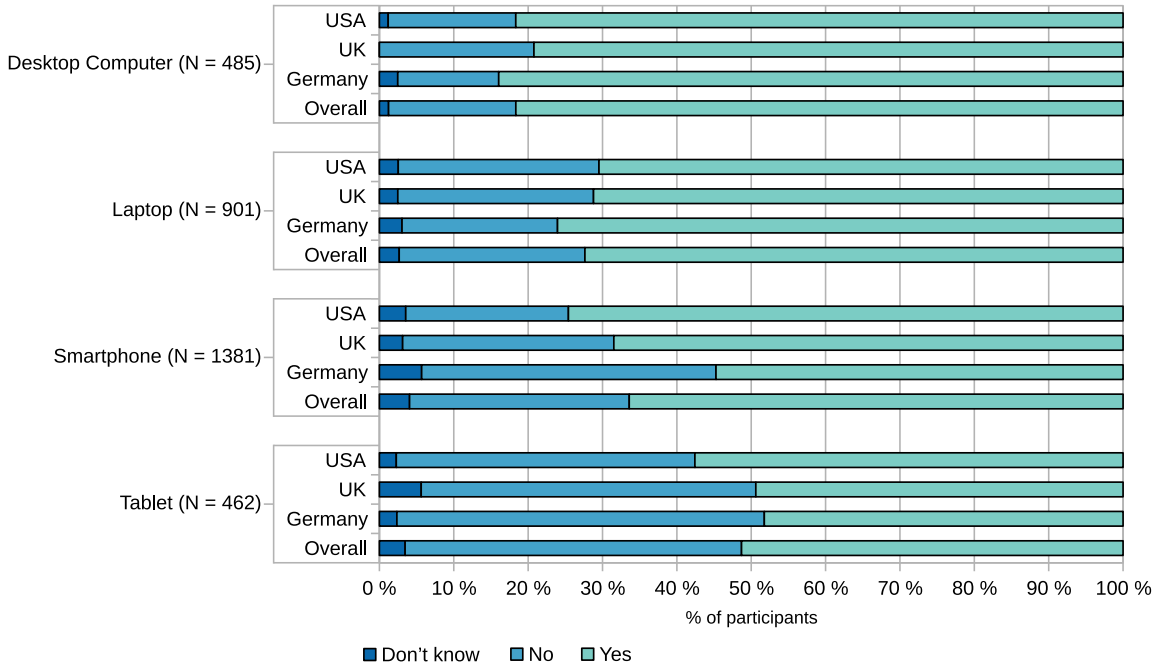


Fig. 5. Use of backup methods for concerned devices, separated by country. The participants had to choose only desktop or only laptop if they had both, depending on which is used more for personal use.

Most participants who did not know whether their devices were backed up or not said they had never considered backups before. Some also indicated that they relied on others to take care of their device, usually without paying for this. Practices of people who make backup are presented in the following Section 5.1. We analyze doing full versus partial backups in Section 5.2 and influence factors on doing backups in Section 5.3. The reasons for not making backups are shown in Section 5.4. We further report participants' attitudes towards backups in Section 5.5, and usage of cloud backups and attitudes to them in Section 5.6.

## 5.1 Making Backups

*Backup Methods and Frequency.* We asked each person which backup methods they use. Participants were free to chose more than one method of backing up. Table 6 shows the percentages of people who used each backup method. In this table, the denominator is always the number of people who back up each device type at all, which is expressed in the row labeled 'N'. For desktop and laptop computers, cloud backups and backups to external hard drives were by far the most popular, with external drives slightly more popular on desktops and cloud slightly more popular on laptops. All of the other methods of backing up were used by less than a third of people who back up desktops or laptops. Builtin backups are more popular for desktop computers than for laptops. Almost one fourth of the respondents use email attachments to backup data on their non-mobile devices. For smartphones and tablets, by far the most popular backup method was cloud backups, with over 80% of people who back up choosing to use the cloud for their backups. Backups to another device are used by one fourth of the smartphone users. Third party tools, network storage and CD/DVD are used very sparsely for all device types. Usage of cloud backups usually means backing up quite frequently, every few

Table 6. Of the people who use backups on each device type, the percentage of those who use each backup method; NA indicates that the question was not asked for some device type.

	Desktop	Laptop	Smartphone	Tablet
Cloud	58.6%	69.8%	83.9%	80.6%
External Drive	73.5%	62.6%	17.6%	17.7%
Builtin Backups	32.1%	23.8%	NA	NA
Backup Software for Mobile Devices	NA	NA	9.9%	8.9%
Email Attachment	23.5%	24.8%	13.2%	10.1%
To Another Device	17.2%	18.1%	25.8%	15.6%
CD / DVD	10.6%	3.7%	NA	NA
Third Party Tool	9.3%	4.1%	NA	NA
Network Storage	7.8%	5.5%	2.1%	2.1%
<i>N</i>	396	652	917	237

days to every time a file is changed. In contrast, backups to external hard drives, other devices or using builtin features<sup>7</sup> take place every few weeks to months.

*Backup Costs.* About 41% of the participants stated that they did not spend money on making backups, whereas 35% of the participants bought additional hardware, and 35% paid for cloud space. Only very small percentage of users bought backup programs (3%) or paid for a backup service (2%), where someone else (a person or a company) takes care of their device and backups. Participants could choose multiple possibilities.

*Reasons for Starting Backups.* Most people (34%) started with backups because they had experienced an adverse incident in the past. Many people began backing up because their system informed them of backup-related features (33%). Around 28% of participants started backing up because relatives, family and/or friends recommended it. Other reasons to commence making backups were discovering the device's backup functions (16%), media information (13%), work requirements (9%), service provider recommendations (5%) and attending training (4%). Participants could select several reasons. Another reason for starting backups or changing the backup strategy might have been the COVID-19 pandemic, as many people started working from home then. We assumed that during this time internal company documents might be used on otherwise privately used PCs and that this could be accompanied by predefined company policies, such as guidelines on backups. However, this assumption could not be confirmed by the study data, as 93% of the participants who worked from home ( $n = 914$ ) stated that their backup behavior had not changed as a result of the pandemic.

*Attitude.* We gave the participants various statements on their backup methods (see Figure 6). Over 80% of participants agreed with most of the statements, for example, that they are satisfied with their backup methods (84%), backed up all relevant files and that they consider the costs reasonable (85%). Most abstentions were on the statements that the backups are secure against unauthorized access, viruses and hackers (over 20%), and most disagreement (around 10%) was on backups being up-to-date and having enough storage.

<sup>7</sup>Built-in features are backup functions offered by the device itself. They can run passively in the background or be actively started by a user.

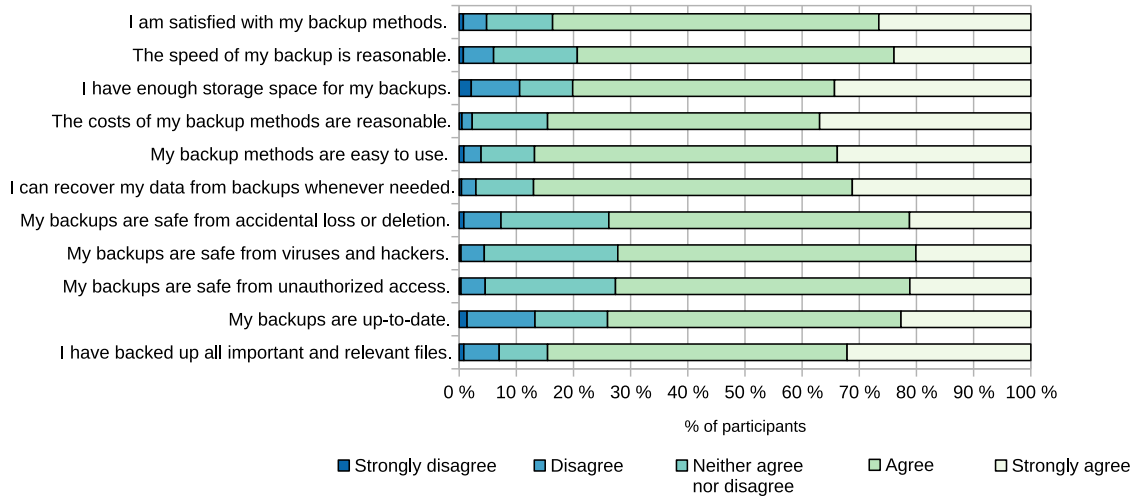


Fig. 6. Agreement on statements concerning backup strategies (N = 1229).

Table 7. Of the people who do backups for each device type, what percentages of respondents chose to do a full vs. partial backup, the percentages within the rows refer to the percentages of the respective device.

	Full Backup		Partial Backup		None	
	Content	Functionality	Content	Functionality	Content	Functionality
Desktop (N = 395)	46.8%	37.5%	50.4%	38.7%	2.8%	23.8%
Laptop (N = 650)	53.1%	34.8%	43.2%	40.3%	3.7%	25.0%
Smartphone (N = 916)	53.5%	46.5%	41.2%	33.8%	5.3%	19.7%
Tablet (N = 237)	54.9%	49.0%	41.8%	30.3%	3.4%	20.7%

*Problems.* Only 54% of the participants stated that they had not had any problems with their backup so far. Frequently mentioned problems were too little storage space (21%), incomplete backups (13%), that the backup (14%) or recovery (9%) process failed. About 22% of participants reported that they do not ensure that their backups were working properly.

## 5.2 Full versus Partial Backups

There was diversity in what people chose to back up. We asked respondents whether they backed up their content data and/or their functionality data. The latter was described as “operating system, programs, settings” for desktop computer and laptop, and as “apps, settings” for smartphone and tablet. We also asked whether they did full backups or only partial backups of each. Table 7 shows their answers broken down by device type. Full user data backups were at similar levels across devices, with the exception of a lower backup rate for desktops. Functionality data – executable files and system configuration – in general was less likely to be backed up than user data. However, functionality data was noticeably more likely to be backed up on smartphones and tablets than on desktops or laptops.

Table 8 also shows the intersection of content and functionality backup. Overall, about 59% of respondents indicated that they do a full backup of their content data on at least one device. About 49% of respondents indicated that they do

Table 8. Of the people who said they do backups, the percentage of people who simultaneously backup content and/or functionality data for at least one of their devices.

		Functionality		
		Full	Partial	None
Content	Full	44%	10%	5%
	Partial	5%	22%	10%
	None	0%	1%	2%

Table 9. Logistic Regressions with the dependent binary variable “doing backup: yes” / “doing backup: no &amp; don’t know”, separated by device type; \*\*\* p&lt;0.001; \*\* p&lt;0.01; \* p&lt;0.05.

	Desktop	Laptop	Smartphone	Tablet
(Intercept)	0.17	-0.52	1.63 ***	-0.19
Prior Experience with Loss	0.80 **	0.89 ***	0.32 *	-0.07
Income	0.06	0.10 *	0.11 ***	0.01
Long-term Partner	-0.48	-0.16	0.17	0.62 *
Others in Household	-0.76 *	0.30	0.32 *	0.75 *
Others in Household, besides partner	1.31 *	0.11	-0.07	-0.28
Has Children	-0.44	-0.35	0.24	0.00
Has Children in Household	-0.23	0.07	-0.10	-0.05
Age (years)	0.02 *	0.02 *	-0.04 ***	-0.00
Gender: Male	0.03	-0.02	-0.45 **	-0.31
Gender: Diverse	0.13	0.72	0.09	14.13
Gender: Prefer not to say	0.26	13.77	-0.63	0.03
N	479	877	1325	446
% Backup	82.7%	74.3%	69.2%	53.1%

a full backup of data that is needed for functionality on at least one device. One popular strategy was making a full backup of both content and functionality data; 44% of respondents indicated that they use that strategy.

### 5.3 Influences on Backups by Device Type

We looked at the backups for each of the four major device types: desktop, laptop, smartphone, and tablet. There were few differences between countries, so all three countries were combined in this analysis. For each device type, we ran a logistic regression. The dependent variable was a binary variable indicating whether the participant said that yes, they did back up their device of that type. Table 9 shows the results of these four regressions.

*Age and Prior Experience with Loss.* First, experiencing a loss seems really important for backing up on a desktop or a laptop (statistically significant and rather big estimates in the same direction). For phones, this is also positive and statistically significant, but it is a much smaller estimate, meaning that it is important, but less so. For tablets, having experienced a loss has almost no effect on backing up.

Looking at age paints a curious picture. Older participants are more likely to backup a desktop or a laptop, but less likely to backup a phone (positive estimates in the desktop and laptop regressions; negative estimate in the smartphone

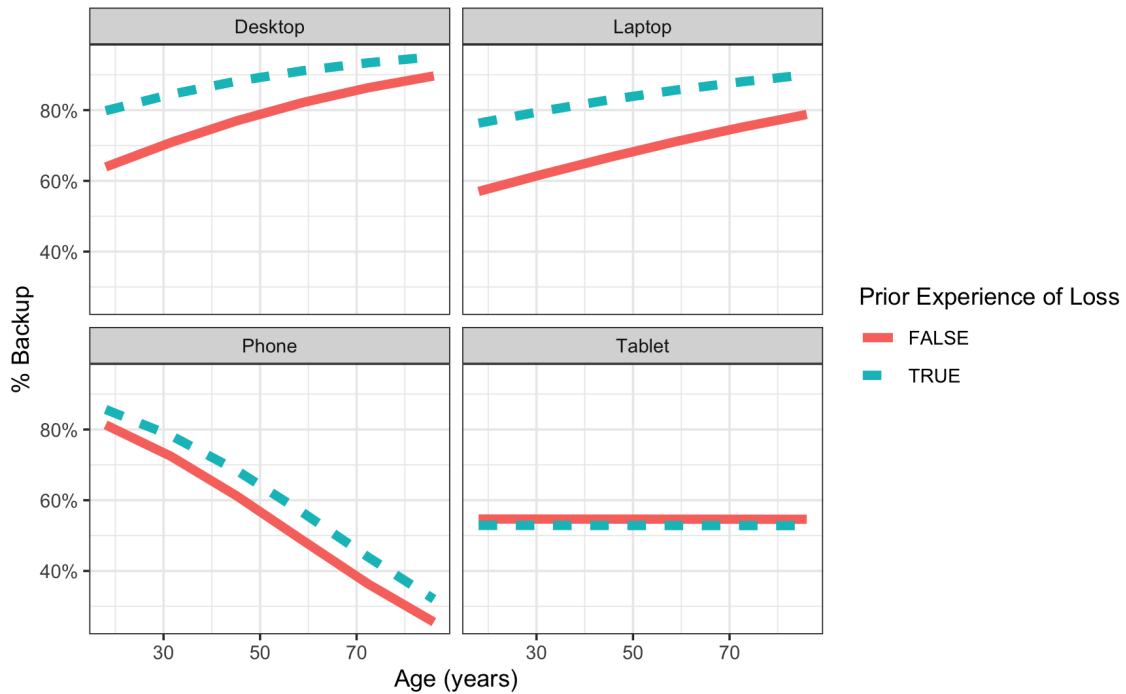


Fig. 7. Predicted probabilities of backing up for Prior Experience with Loss and Age, separated by device type.

regression). All three are statistically significant; the magnitude is about twice as large for the negative backing up of phones than the positive effect of backing up desktops and laptops. Furthermore, there is no effect for tablets.

Figure 7 shows predicted probabilities from these models graphically. The effects of age can pretty clearly be seen on the graphs. For desktops and laptops, age has a big effect, with the probability of backup increasing with age. For desktop, an 18 year old man<sup>8</sup> has a 62% chance of backing up their desktop. This increases to 75% for a 45 year old, and to 84% for a 72 year old. For laptops, an 18 year old man has a 58% chance of backing up. This goes up to 68% for a 45 year old, and 77% for a 72 year old. For phones, the effect is quite the opposite, and even more striking. An 18 year old man has a 80% chance of backing up their phone. This goes down to 60% for a 45 year old, and 37% for a 72 year old. The difference between people who experienced loss and who did not is smaller, and the lines drop vertically a lot more, meaning that there is more effect of age, and less of an effect of prior experience with loss.

*Income.* We measured income using a standard scale of 11 income categories. In general, people with higher incomes are more likely to back up their data. This is especially true for smartphones, where individuals in the highest income category are almost 20% more likely to back up their smartphones than individuals in the lowest income category. This

<sup>8</sup>To understand the estimates better, we use predicted probabilities and graphs of those probabilities to interpret regressions, instead of relying solely on regression coefficients like Table 9. To generate an estimated probability from a logistic regression like the one we consider here, a value must be chosen for *all* of the independent variables and the model is used to calculate the probability of backup for a hypothetical person with those values. Most variables have a logical “null” value (i.e. NO to the question), but as there is no logical null value for gender, we had to pick one gender. For these estimates, there is a difference between men and women, it is analyzed in a separate paragraph later in this section. Since in this paragraph we are trying to interpret the effects for age and for prior experience with loss, it doesn’t matter whether we choose a man or a woman, but we need to choose one. And to understand the comparison, it helps to make the same choice for all of the estimated probabilities (i.e., hold everything constant except the variable of interest).

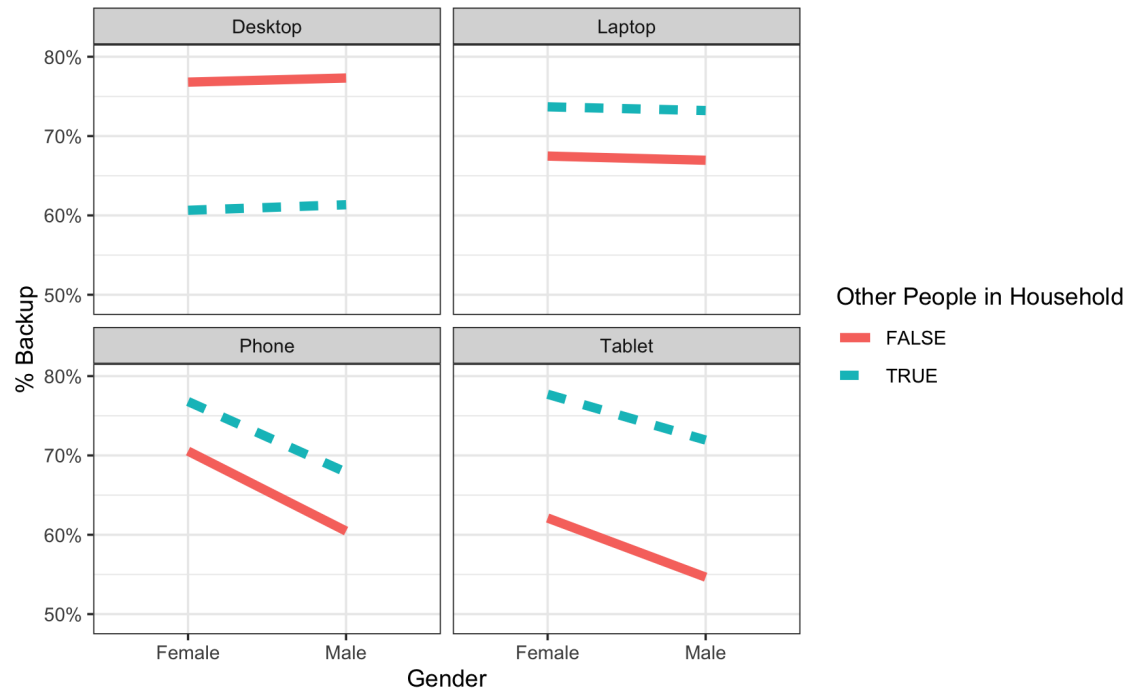


Fig. 8. Predicted probabilities of backing up for gender and others in household.

is a statistically significant effect for smartphones. A similar pattern exists for desktops and laptops, though not quite as big of an effect and is not statistically significant for desktops. There were almost no differences by income for whether tablet users back up.

*Gender and Other People in Household.* Additional variables play a (statistically significant) role in determining whether people back up each device type. A person with a desktop is more likely to backup if they have other people in the house *other than* their partner. Having other people in the house also has a small positive effect of backups for laptops, but it is not statistically significant. For phones, a participant is more likely to backup if they have other people in the house, *inclusive* of their partner. So, for desktops the partner is not included, but for phones they are in the positive effect.

There is also an interesting gender effect. For desktops and laptops, there is almost no gender effect; men and women were about equally likely to back up them. However, for phones, there is a bigger difference; women were much more likely to backup their phone than men were. There is a similar magnitude (though not statistically significant) difference between women and men for tablets as well; this lower statistical significance could be a result of the smaller number of people (smaller N for the regression) who use tablets.

Figure 8 shows predicted probabilities by varying gender and whether the respondent has others in their household. A person has about a 76% chance of backing up a desktop if there are other people in the house, and a 60% chance if there are no others in the house. For laptops, the opposite is true – 67% vs 74%. Living alone seems to really matter for tablets: a woman living alone has about a 61% chance of backing up her tablet; the same woman living with others has

Table 10. Reasons for not backing up concerned devices in percentage, multiple choice.

	<b>Desktop Computer</b> N = 83	<b>Laptop</b> N = 225	<b>Smart- Phone</b> N = 408	<b>Tablet</b> N = 209
I have never thought about backups.	28.9%	32.0%	35.5%	25.4%
I don't know how to backup.	16.9%	27.1%	27.2%	15.3%
I haven't tried doing backups, because I think it is too complicated.	18.1%	20.9%	18.9%	14.4%
I tried to do backups, but did not succeed.	2.4%	5.8%	4.2%	2.4%
I used to do backups in the past, but stopped.	22.9%	18.2%	10.8%	8.1%
I don't have time for backups.	18.1%	16.9%	12.7%	11.0%
I don't have the money for backups, it's too expensive.	14.5%	10.7%	9.3%	7.7%
I don't need backups, as I have not had any problems so far.	36.1%	26.7%	23.8%	20.6%
I don't need backups, as nothing is going to happen to my device.	8.4%	7.1%	6.9%	8.1%
I don't need backups, as I have nothing important on this device.	32.5%	26.7%	34.6%	52.2%
Someone else is taking care of backups for my device.	7.2%	7.6%	4.9%	5.3%
Other reasons	3.6%	8.4%	6.4%	3.3%

about a 74% chance. Men and women are about equally likely to back up their desktops and laptops (horizontal lines in the graphs in Figure 8). However, women are more likely than men to back up their phones and tables. For both phones and tablets, this difference is about 10% (e.g., 70% for women living alone vs. 60% for men living alone).

*Tablets.* Backup behavior for tablets looks different than the other device types. Neither prior experience with loss nor age has much effect on whether someone backs up a tablet – both of which are very important for the other three device type. The only variables that seem to predict whether someone will back up a tablet is whether they have a partner, and separately whether they have other people living in the home. Both have positive, medium-sized effects. The gender difference is similar to phones – women are more likely to backup tablets than men.

#### 5.4 Reasons for Not Making Backups

Table 10 presents the reasons for not creating backups (multiple choice). A common reason is that the participants have never thought about backups before. They often do not consider backups to be necessary for the device in question, because they have not stored anything important on it. This reason is especially popular for tablets, with over 52% of users saying so. Many also have not experienced any problems with the device so far. P416 stated: *“I’ve never really gave it much thought. I don’t really need a backup method as I don’t really have anything important to worry about.”*

Quite a high number of participants (14.4% to 20.9%, depending on device) feel that the backup process is too complicated. Some also indicate that they don't have enough time or money to backup their data. Others also indicated that they had carried out backups in the past, but had stopped doing so. P200 stated: *“The services I used to use for backing up files became too expensive for me, so I just don’t keep anything on my devices that would be difficult to replace were my desktop to fail,”* and P175 mentioned: *“Was paying additional costs for extra storage and when phone crashed, nothing was backed up on Cloud with extra storage I had been paying for years”.*

Some participants stated that they did not know how to create backups, particularly for laptops and smartphones, as P453 summarizes: *“I’m not sure. I did everything the device asked but it wouldn’t backup.”*

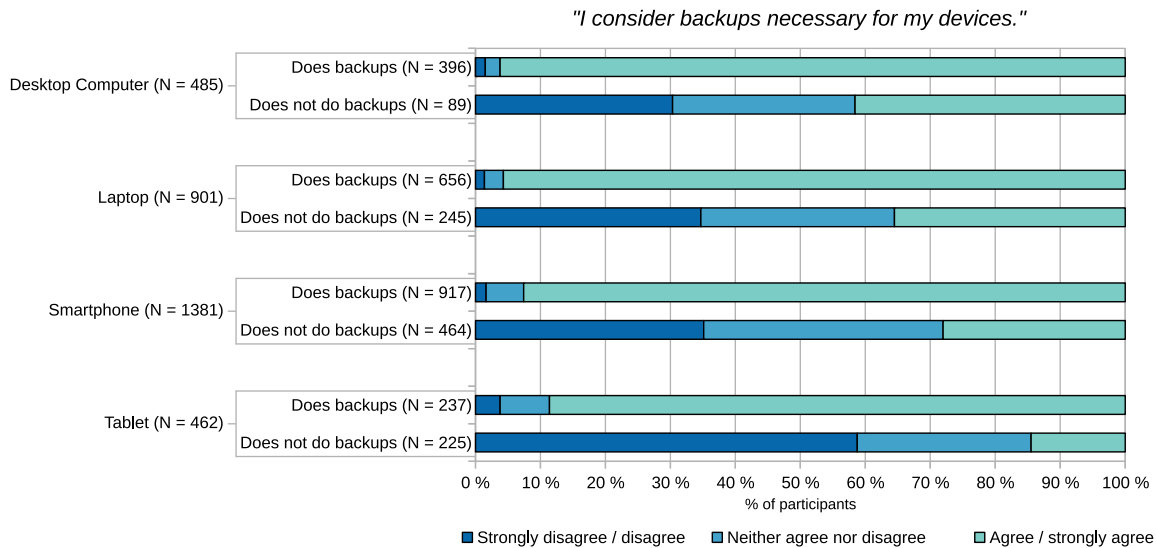


Fig. 9. Agreement with the statement *"I consider backups necessary for my devices."* in comparison to whether participants backup their devices or not.

## 5.5 Attitude to Backups

Although some devices are not backed up, their users consider the backups necessary for their device (see Figure 9). Specifically, 30-40% of participants who do not back up their desktop computers, laptops and smartphones agree that creating backups is essential. This indicates that people understand the importance of backing up their data, but there are still barriers to actually creating them.

We asked all participants in a free text question to name characteristics that are important to them regarding a backup method, independently on whether they do backups or not (see Table 13 in Appendix B for assigned codes). The most often mentioned characteristic was that the backup should be easy to use (51%) and that the backup should reliably save and restore data (28%). For many participants, it is also important that the backup and restore is fast (23%) and that the backup is safe from hackers or viruses (19%). There was no difference between people who do backups, and those who do not.

The costs also play an important role, which should be appropriate to the backup functions (14%). Some participants mentioned automated backups (12%), for example by using the cloud. P918 writes here: *"Since it has become a practical and affordable option I have relied entirely on automatic backup to the cloud of all my user files as and when they change. As a former IT professional I am aware that this effectively outsources to my service provider the responsibility for really doing backups properly at the data center level. I hope that they are actually doing this!"*.

There is also a focus on ensuring that the data is accessible (10%). P324 summarizes this: *"That it is easy to do. That files are easy to access afterward. Speed isn't as important if these two things are in place. I'd rather backups take time and be easy to access than have them be very fast but not easily accessible"*. On the contrary, for others speed is important, as P145 summarizes: *"I want to be able to back up my important files but not have to deal with the slowness and large space requirements of backing up the entire system"*.

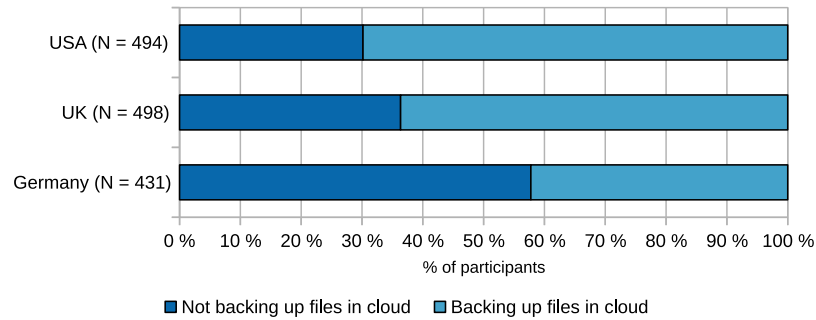


Fig. 10. Overview of how many participants backup their files using cloud services, separated by country.

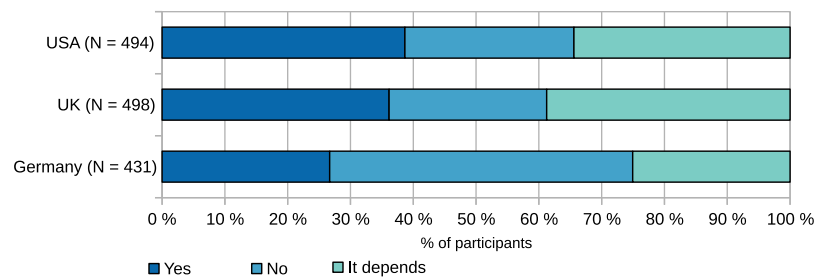


Fig. 11. Agreement if participants would like to backup all their data, programs and settings in the cloud, separated by country.

## 5.6 Cloud Backups: Usage and Opinions

The results previously showed that the cloud is a widely used backup method. Overall, 77% of participants stated that they use cloud services to backup their data. The most commonly used cloud providers are Google Drive (62%), Apple iCloud (43%), Microsoft OneDrive (31%) and Dropbox (28%). The cloud is also utilized to access data from different devices (61%), to save storage space (49%) or to share data with others (37%). Whereas 43% of participants think that the cloud provider creates full backups of their data, 16% think that at least partial backups are made, while 35% do not know whether the backups are created or not. There is a difference between Germany and two other countries, with cloud backup being much less popular in Germany, see Figure 10.

Although cloud is associated with security risks, we thought that it also offers an easy and intuitive form of backup for content, and also for functionality of mobile devices. Therefore, we were curious whether participants would like to store all their content and functionality data in a cloud. The results, again, differed between Germany, where 48% said 'No', and the two other countries, where around 26% said 'No', see Figure 11.

We requested participants justify their decision as a free text response (see Table 14 in Appendix D for codes). Those who would like to back up everything explained that they felt comfortable having a backup of their data in this way. It is also easier to access the data from different devices or to restore the old state on a new device if necessary. According to those participants, the cloud is very easy to use and also serves as an outsourcing of storage space. P105 summarized this: *"I can easily access, backup, and restore from anywhere"*.

Participants who did not want to store all their data in the cloud often justified this by saying that they would only upload certain data to the cloud, because they did not trust it. P73 stated: *"I don't feel comfortable doing all my data and*

stuff on the cloud because it is not free from security risks. If it is sensitive information I back it up on an external device that only I have access to. P1415 mentioned: “I feel safer when I have control over my backups; I don’t have that level of trust with the cloud.” Fear of unauthorized access, such as by hackers, as well as privacy concerns, is widespread among these participants. German participants expressed general distrust and explicit privacy concerns more often (21.3% and 8.8%, respectively), as compared to the UK (12.4% and 4.0%) and the USA (9.5% and 6.1%).

Another argument against cloud storage is the limited storage space and the associated costs if extra storage space is requested. P750 “would consider it if cost was lower”. Other participants justified their decision by saying that their data was not important enough to backup. P639 stated: “I don’t store anything on my devices that needs to be backed up.”

The participants who answered ‘It depends’ justified that the cloud should be secure and easy to use, as well as being affordable. P45 mentioned: “I might use it if it was more convenient, reliable, and inexpensive. The ones I’ve checked out in the past weren’t, so I use other methods.” Some participants also stated that they did not know exactly how the cloud works. P543 stated: “I would if I knew how. They don’t teach older people how to use them”.

## 6 DISCUSSION

Almost half of participants in our study reported experiencing at least one data loss incident. These data loss incidents happened across all major device types. This shows that data loss is surprisingly common, and something that people regularly need to deal with. There is a clear need for resilience against data loss.

No matter whether that loss is caused maliciously or accidentally, a good way to deal with such an incident is by backing up data. Indeed, “make backups” is common advice from cybersecurity experts toward regular computing users. Our results, however, highlight two major challenges that show that advice is underspecified and insufficient: that the usability of recovery is a major barrier and stress point for many users, and that confusion about the need to back up both, content and functionality, limits the effectiveness of backups.

### 6.1 Usability of Recovery and the Role of Backups (RQ1, RQ2)

When a data loss incident occurs, it is useful to have a backup from which the data can be recovered. However, even when people have backups, the results of our study indicate they are not always able to recover their data. Thus, despite the presence of backups, many of our participants had a lack of resilience.

Recovering data from a backup is not nearly as straightforward as it may seem, as a substantial number of participants described many different types of problems with recovery, such as that backups were outdated or incomplete or that the recovery process is overly complicated. While it seems to be more difficult to recover from data loss without a backup, it is still not easy or straightforward even if there is a backup.

Furthermore, most participants described the recovery process as strenuous and difficult. The highest levels of stress were felt by people who made backups, but encountered problems with them during recovery. We can support the observation by Kljun et al. [13, p. 13]: “Restoring information is a daunting process if there is no backup or if backup fails.” They asked their participants to describe a story that changed how backup is done. Although not many details are presented in the paper, backup and recovery processes described by their participants do not seem highly usable, because users often misunderstood how technology works or made other mistakes. Recovery usually occurs at a time with heightened stress (during a data loss incident); it is not good if the recovery process itself adds stress and challenge to this already stressful time.

Our participants seemed mostly satisfied with their backup methods. The backup part of backup-and-recovery seems to be good for a high percentage of users. However, the recovery part causes stress and anxiety, and often does not work

to fully recover. Our data suggests that additional attention needs to be paid to the usability of the recovery process, to make the recovery process both more successful and less stressful. If people are to be resilient against data loss, they need to be able to recover and not just back up their data.

One could argue that some participants might have made a conscious trade-off: Because backing up is not easy and requires expertise and time, they only backed up the most valuable data that cannot be recovered, and were prepared to spend time and nerves to recover the rest without backups, because recovery incidents are rare. We cannot say in how many cases this happened, and how many of those participants regretted that their backups were incomplete once they were in the recovery situation. This would be an interesting avenue for future work (Section 6.7).

## 6.2 Content versus Functionality (RQ2, RQ3)

Resilient computing means being able to get the computing device back to the state where it was before a data loss incident, so that the user can keep using it. It is not enough to just get the content back; being resilient means also being able to use the device as before. Our findings suggest that there are at least two different types of data that are important to back up: content and functionality. Content data may comprise either irreplaceable personal content (documents, photos, home movies) or costly third-party content (music, movies). Functionality data, on the other hand, is files that are essential to keep a computing device functional; this includes both the files that run the system (executable files, system files, apps) and the individualized configuration details.

It seems that participants think about these two types of data differently. Most people seem to think in terms of content loss, not functionality loss, and were more likely to back up content than functionality data. They were also more likely to do full backups of content than of data for functionality. Kljun et al. [13] found the same tendency 10 years ago (their study was run in 2013). This suggests that people are generally more prepared for content losses than functionality losses. However, in many of the data loss incidents that they experienced, loss of functionality was one of the major issues that arose, and one that was less likely to be resolved successfully. Indeed, it was often lost functionality that caused people to first notice the data loss incident.

The recovery process seemed to be more challenging for functionality losses than for content losses. The most common recovery tactic – restoring data from backup – works well for both content and functionality data for some of our participants. However, many of the other ways that people tried to recover are primarily intended to address functionality losses such as buying a new device, attempting to reboot or running anti-virus software. These are all attempts to restore functionality, not content.

According to these findings, data loss incidents frequently require extra efforts to recover functionality. However, most people focus their backups on content, not on functionality. Being resilient should enable people to both easily recover their content, and easily get their devices back to a working state. However, most of the advice from experts isn't nuanced enough to make this distinction; instead, it just emphasizes the importance of making backups. We recommend that advice about backups include that both content and functionality data be backed up, and that recovery should be improved to make it easier to recover functionality in addition to content.

## 6.3 Cloud for Backup? (RQ3)

Although most participants use the cloud to make backups, its usage is different per device, and also per country. German users, in particular, exhibit a distrust in cloud providers, citing privacy violations. However, participants from all three countries also mentioned other disadvantages of cloud backups: lacking control and uncertainty about availability of data to support recovery if required. These disadvantages were also mentioned in the study by Muslukhov et al. [16],

who surveyed participants at a university. Kljun et al. [13] showed that users often feel skeptical about the cloud, as they believe that others may also have access to their data.

This demonstrates that even in developed countries such as those we studied, not everyone has a continuous Internet connection with enough bandwidth to be able to back up all their content and functionality data into the cloud. Moreover, some older participants indicated that they did not know how to use the cloud for backups. We note that these participants were recruited through online crowdworker platforms, and could reasonably be expected to be comfortable with the online world. This calls for better explanations about how cloud backups work, and its advantages and disadvantages.

Menard et al. [14] surveyed intention to use cloud backups among students at a US university. The most important factor was convenience of this backup method, followed by high risk perception of data loss. We found a similar effect when we asked whether the participants would like to use cloud backup for all their data and functionality, as participants who said ‘yes’ gave convenience and fear of data loss as their main reasons.

Even so, local backup methods are still important. In Muslukhov et al. [16]’s study, participants stated that they would prefer to use local backups rather than cloud backups. It is unfortunate that not many people use networked storage for backups. Combining the advantages of cloud and local backups creates a kind of a local cloud that is under the user’s full control. However, this comes at extra cost for hardware, and needs installation and maintenance, which requires non-trivial expertise. These might be the reasons for low usage of networked storage.

#### 6.4 Who Backs Up? (RQ4)

Our findings also suggest that resilience is unevenly distributed among users. Some people are more able to be resilient against data loss than others. One of the main influences on resilience is prior experience. Users who have previously experienced data loss are much more likely to back up their data, and thus be able to recover. This was also shown in the work of Menard et al. [14], although here only young students were asked about intention to use cloud backup in theoretical scenarios. Considering our study, this is unsurprising, but also unfortunate. As a large number of our participants have experienced data loss, we should not rely on waiting until after they have a negative experience to help them with resilience. Instead, we should find better ways to proactively encourage backing up and recovery.

We found a surprising difference in backups by age and device. Older adults are more likely than younger adults to back up their desktops and/or laptops, whereas younger adults were more likely than older adults to back up their smartphones. It is not clear why this pattern exists. It is possible that older adults are more likely to use their desktop or laptop as their primary computing device, and younger adults are more likely to use a phone as their primary computing device. On the other hand, it could be that older adults are more familiar with the backup technologies that work well with desktops / laptops, which are different than the technologies that work with phones.

We also found evidence that women were much more likely to backup their phone than men. The study by Redmiles and Hargittai [25] showed similar results for US students.

Further, we found an inequality related to income. Users with higher incomes are generally more likely to back up their data. Backups cost time and money, so in many ways it makes sense that people with higher incomes might be more likely to back up. It is also possible that they use their computing devices for work and financial transactions more than lower income people, which might increase their perception of the necessity of resilience. This finding highlights a potential inequality, that lower income people are less likely to have backups in the case of a data loss incident and thus are less resilient.

## 6.5 Implications

*Educational Implications.* Experts advise users to back up their data [29]. However, our results indicated that not all participants backed up all of their important data. Most started backing up after losing data forever due to an incident. This highlights the need for better education on the importance of backups to ensure data resilience. In particular, misconceptions about backups must be addressed. Our findings revealed that users primarily focus on content backups, while functionality backups are often overlooked. For full resilience, it is crucial to perform both content and functionality backups. Even if there are trade-offs between making complete backups and resources needed for recovery, users need more information to fully understand these trade-offs and make informed decisions.

Many participants faced challenges when trying to restore data from their existing backups. This highlights the necessity for educational resources that tackle common issues related to backups, such as handling outdated or incomplete data and recognizing which data needs to be included in the backup. It is also important to provide guidance on the process of restoring data from a backup.

*Design Implications.* Even the best educational materials have their limits if the backup and recovery mechanisms are hard to use, even if users want to do backups.

We saw that there are many challenges that users face when doing backups. Although we are unable to point out which particular tasks of which particular backup and recovery solutions go wrong, as we did not investigate this, our study makes clear that there are usability and user acceptance impediments to backup.

To make the backup and recovery process as accessible as possible, one design implication is to enhance the usability of backup creation. This means designing the process so that it is not time-consuming or cumbersome for users. Based on our data, we cannot say exactly how the process should be designed or what design factors should be included. A future study could address this, as we suggest in Section 6.7.

Ideally, backups should be able to run automatically in the background, enabling users to continue with their main activities without interruption. Devices should offer easy-to-use backup functions as standard, so that there is no need to purchase third-party software. Ideally, the devices should then regularly remind the user to perform a backup if an automatic backup has not already been activated. Moreover, cost considerations are crucial; backup methods should remain affordable, ensuring that a wide range of users can implement them without financial strain. Backup methods should also provide enough storage space to avoid the risk of an incomplete backup.

A robust backup solution is vital as it not only protects against data loss, but also provides users with greater confidence in their ability to recover from potential incidents. Thus, reliable backup methods should provide better support in preparing for potential incidents, enabling users to effectively recover all lost data. It should also be possible to verify whether data can be completely restored from the backup in the event of an incident. Overall, a well-designed backup and recovery process can significantly reduce stress and improve productivity for users.

## 6.6 Limitations

Our results are limited to the USA, UK and Germany, which are all WEIRD (Western, Educated, Industrialized, Rich and Democratic) countries with similar, although not equal, cultures. A study with participants from other countries, especially those with different cultural backgrounds, infrastructure, wealth and political situations, could provide different results with regard to backup and recovery behavior and experiences.

We asked participants to tell a story about the recovery incident they remembered most. In this, we followed Rader et al. [24], but also our own wish was to know about the incident that was most remembered, and thus, probably

influenced the participants most. Therefore, our data on recovery incidents are biased towards extremes. We also cannot ensure that participants correctly recalled the incidents in which data were lost.

Furthermore, the participants were recruited via Prolific and Clickworker crowdworker platforms. Participants on these platforms may be different from the general population, even in the samples that are representative of the particular country in age and gender, which limits the generalizability of their answers. For example, Tang et al. [34] found that their representative in age, gender and ethnicity US-based sample on Prolific was higher educated than the general population, which we also found for our Prolific and Clickworker samples [31, 32]. In addition, we had fewer German participants than the UK and USA samples. Although several attempts were made to recruit 89 German males over 57 years old and 106 German females of this age, only 72 and 48, respectively, could be recruited on the Clickworker platform.

The questionnaire might be considered quite long with an average completion time of 15 minutes. We therefore added two attention tests to ensure that the effect of survey fatigue was reduced. Participants who did not pass the attention tests were still compensated, but their data was not included in the analysis.

We decided to word the statements in the survey the participants could agree or disagree with only in one way (negative for recovery attitude, positive for backup attitude). We followed the recommendation by Suárez-Álvarez et al. [33] to word the statements in only one direction to avoid putting additional stress on the participants in a long questionnaire. Furthermore, when mixing positive and negative statements, the negative statements must be reverse-coded in the evaluation. This is often debated in related work, as linguistically, agreeing with a positive statement is not the same as disagreeing with a negative statement [6, 37]. Additionally, the statement “I want to backup all my data in the cloud” could have been interpreted in two ways. We intended to ask about the attitude to full cloud backups. However, participants who use local backups might have answered “No” to this question, as they do not need the cloud for backups.

## 6.7 Future Work

The exploratory nature of this study allowed us to uncover and describe trends in backup and recovery behavior, but it did not enable us to fully understand the reasons behind these trends. This leaves room for future research. Furthermore, only the most general trends are presented in this paper, leaving other discovered trends to future work.

*Differences in Backup Behavior.* Our results indicate that there are differences between the three countries we studied: the USA, the UK and Germany. We believe that they may be attributed to differences in cultures (e.g., beliefs and attitudes), experiences, demographic factors, and the varying infrastructures of each country. Future research could focus on the underlying causes of these disparities, and investigate backup and recovery behavior in other countries.

We found that individuals with higher income are more likely to backup their data, especially their laptops and smartphones. One assumption behind this trend may be the relationship between income and the choice of operating systems. For example, it could be that higher-income individuals are more inclined to own iOS phones rather than Android ones. For iOS phones, there are many built-in backup features that might be more usable than features that Android phones offer. It would be valuable to explore not only the preferences for specific devices and backup methods, but also to understand the motivations behind these choices.

In general, there are differences in backup behavior between different operating systems, and between the 16% of participants with computer science background versus participants without this background. There also differences between devices that are mostly utilized for personal use versus devices that are also used for work. We were unable to

report all these differences in the present paper, because we concentrated on most important trends, and leave these investigations to future work. Reasons behind these differences should be investigated in follow-up studies.

*Usability of Backup and Recovery.* Our study revealed that users frequently encounter various barriers when trying to utilize backup solutions effectively. Additionally, 15-40% of participants, depending on the device, consider backups necessary, although they do not make them. We do not know the reasons behind this behavior, but it could well be due to inadequate usability. To address these issues, future research could investigate the usability of different backup solutions, including the recovery process. This could help to develop strategies to enhance the accessibility of backup methods, making them easier for users to implement and manage. Another future study could also look at the design of backup methods and determine which design factors would improve overall usability. Concrete design concepts should also be explored.

*Nuances and Trade-Offs in Backup Behavior.* We observed that certain types of data appear to be more valuable to users than others, e.g., content data is backed up more than data that ensure functionality of the device. Another finding is that many users opt for partial backups of data and functionality, or backup some, but not all of their devices. An explanation could be that individuals may not see backups as necessary for a specific device if the data on that device is already stored on another device. Moreover, a lot of data can be reacquired if data loss occurs, for example through download or purchase. This raises questions about how users prioritize what to backup. In addition, the backup process could also be situation-dependent and influenced, e.g., by the perceived stress level or risk. Future studies could explore how the perceived importance of specific data types or situational factors affect users' decisions to perform complete backups of their devices, as opposed to opting for partial backups or no backups.

Additionally, it would be valuable to investigate the economic trade-offs that users consider when deciding on backup options. This includes understanding what costs (including time and effort) they deem acceptable for a comprehensive backup method and how these costs impact their overall decision-making process. We examined the experiences of users during the recovery process after data loss. Many participants started backing up their data or adopted multiple backup methods afterwards. Although the recovery process seemed stressful, we cannot determine whether the perceived stress was appropriate for the participants in retrospect, nor can we identify how many individuals regretted that their backups were incomplete once they were in the recovery situation. Future research could explore these issues further.

## 7 CONCLUSION

We set out to explore the backup-enabled resilience of devices for personal use in three countries. Previous studies had revealed low engagement with backing up [13, 25], but we discovered very high levels of engagement with making backups of some kind. Our participants told revealing stories about their experiences recovering from adverse incidents, both with and without backups. Our findings deliver lessons for awareness raising efforts: there is a need to include 'what' to backup additionally to ensuring that people know that they ought to make backups. Moreover, the role of backups in easing recovery and enabling resilience should be highlighted additionally to advising people to make backups. If people only understand that backing up is imperative and are unaware of the role of backups in recovery and resilience, it would be unsurprising that they do not make comprehensive backups of their devices. Although recovery is also possible without backups, and making backups constitutes a trade-off between the completeness of backups and resources needed for recovery, it is not clear whether users engage in these trade-offs with full understanding of consequences of incomplete backups. Furthermore, even if they want to make comprehensive backups, recovery is

undeniably challenging due to the unusable nature of the recovery process. Improving usability would make it possible for users to improve their personal resilience in the face of data loss.

*Acknowledgements.* We thank Freya Gassmann and Victoria Karaseva for supporting us in creating the questionnaire. We thank the anonymous reviewers for their valuable feedback, which further improved the paper. This work was partially funded by the German Federal Ministry of Education and Research under grant 16KIS1271K.

*Author Contribution Statements.* Julia Wunder: Methodology, Formal analysis, Investigation, Data Curation, Visualization, Writing – Original Draft, Writing – Review & Editing. Rick Wash: Formal analysis, Data Curation, Writing – Original Draft, Funding acquisition. Karen Renaud: Resources, Visualization, Writing – Original Draft, Writing – Review & Editing. Daniela Oliveira: Conceptualization, Methodology. Zinaida Benenson: Conceptualization, Methodology, Formal analysis, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

## REFERENCES

- [1] Syed Ishtiaque Ahmed, Shion Guha, Md Rashidujjaman Rifat, Faysal Hossain Shezan, and Nicola Dell. 2016. Privacy in repair: An analysis of the privacy challenges surrounding broken digital artifacts in Bangladesh. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development*. ACM, Ann Arbor, USA, 1–10. <https://doi.org/10.1145/2909609.2909661>.
- [2] Elham Abdullah Al-Qarni. 2023. Cybersecurity in Healthcare: a review of recent attacks and mitigation strategies. *International Journal of Advanced Computer Science and Applications* 14, 5 (2023). <https://doi.org/10.14569/IJACSA.2023.0140513>.
- [3] The World Bank. 2023. The World Bank Atlas method - detailed methodology. <https://datahelpdesk.worldbank.org/knowledgebase/articles/378832-the-world-bank-atlas-method-detailed-methodology> accessed in April 2023.
- [4] Karoline Busse, Julia Schäfer, and Matthew Smith. 2019. Replication: No One Can Hack My Mind. Revisiting a Study on Expert and Non-Expert Security Practices and Advice. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX, Santa Clara, USA, 117–136.
- [5] Ann Chervenak, Vivekenand Vellanki, and Zachary Kurmas. 1998. Protecting file systems: A survey of backup techniques. In *Joint NASA and IEEE Mass Storage Conference*, Vol. 99. Maryland, USA.
- [6] Seung Youn (Yonnie) Chyung, Katherine Roberts, Ieva Swanson, and Andrea Hankinson. 2018. Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale. *Performance Improvement* 56, 10 (2018), 15–23. <https://doi.org/10.1002/pfi.21727>
- [7] Clickworker. 2023. Our Clickworker community. <https://www.clickworker.com/clickworker-crowd/> accessed in Feb 2023.
- [8] Trevor Cooke. 2024. Backup Statistics in 2024: Recovery & Data Loss. <https://earthweb.com/blog/backup-statistics/> Accessed 20/11/24.
- [9] Edelman Trust Institute. 2024. 2024 Edelman Trust Barometer Global Report. Accessed 6/9/24 <https://www.edelman.com/trust/2024/trust-barometer>.
- [10] Federal Office for Information Security. undated. Backups – what’s the best approach? Accessed 6/9/24 <https://www.bsi.bund.de/dok/6598926>.
- [11] Iulia Ion, Rob Reeder, and Sunny Consolvo. 2015. ... No one can hack my mind: Comparing Expert and Non-Expert Security Practices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX, Ottawa, Canada, 327–346.
- [12] Lee Kim. 2022. Cybersecurity: Ensuring confidentiality, integrity, and availability of information. In *Nursing Informatics: A Health Informatics, Interprofessional and Global Perspective*, Ursula H. Hübner, Gabriela Mustata Wilson, Toria Shaw Morawski, and Marion J. Ball (Eds.). Springer, Switzerland, 391–410. [https://doi.org/10.1007/978-3-030-91237-6\\_26](https://doi.org/10.1007/978-3-030-91237-6_26).
- [13] Matjaž Kljun, John Mariani, and Alan Dix. 2016. Toward understanding short-term personal information preservation: A study of backup strategies of end users. *Journal of the Association for Information Science and Technology* 67, 12 (2016), 2947–2963. <https://doi.org/10.1002/asi.23526>Citations:12.
- [14] Philip Menard, Robert Gatlin, and Merrill Warkentin. 2014. Threat Protection and Convenience: Antecedents of Cloud-Based Data Backup. *Journal of Computer Information Systems* 55 (09 2014), 83–91. <https://doi.org/10.1080/08874417.2014.11645743>
- [15] Luka Murn. 2021. Data Safety and Cybersecurity. In *Digital Transformation of the Laboratory: A Practical Guide to the Connected Lab*, Klemen Zupancic, Tea Pavlek, and Jana Erjavec (Eds.). Wiley Online Library, 85–100. <https://doi.org/10.1002/9783527825042.ch4>.
- [16] Ildar Muslukhov, Yazan Boshmaf, Cynthia Kuo, Jonathan Lester, and Konstantin Beznosov. 2012. Understanding Users’ Requirements for Data Protection in Smartphones. In *2012 IEEE 28th International Conference on Data Engineering Workshops*. IEEE, Arlington, USA, 228–235. <https://api.semanticscholar.org/CorpusID:2942993>
- [17] Johannes Nakayama, Nils Plettenberg, Patrick Halbach, Laura Burbach, Martina Ziefle, and André Calero Valdez. 2019. Trust in cyber security recommendations. In *IEEE International Professional Communication Conference (ProComm)*. IEEE, Aachen, Germany, 48–55. <https://doi.org/10.1109/ProComm.2019.00014>.
- [18] National Cyber Security Centre. undated. Backing up your data. Accessed 6/9/24 <https://www.ncsc.gov.uk/collection/top-tips-for-staying-secure-online/always-back-up-your-most-important-data>.
- [19] Statistical Offices of the Federation and Germany the Länder. 2022. Zensus Datenbank - Persons: Sex - Age. <https://ergebnisse.zensus2022.de/datenbank/online/table/1000A-2026> accessed in Feb 2023.

- [20] Suzanne Prior and Karen Renaud. 2023. Who is best placed to support cyber responsabilized UK parents? *Children* 10, 7 (2023), 1130. <https://doi.org/10.3390/children10071130>.
- [21] Suzanne Prior and Karen Renaud. 2024. Are UK parents empowered to act on their cybersecurity education responsibilities?. In *International Conference on Human-Computer Interaction*. Springer, 77–96. [https://doi.org/10.1007/978-3-031-61379-1\\_6](https://doi.org/10.1007/978-3-031-61379-1_6).
- [22] Prolific. 2023. Prolific Audience Checker. <https://app.prolific.com/audience-checker> accessed in Feb 2023.
- [23] Emilee Rader and Rick Wash. accessed in December 2017. Materials for SOUPS 2012 paper “Stories as Informal Lessons About Security”. <https://osf.io/9dfzn/>.
- [24] Emilee Rader, Rick Wash, and Brandon Brooks. 2012. Stories as informal lessons about security. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*. ACM, Washington, USA, 1–17. <https://doi.org/10.1145/2335356.2335364>.
- [25] Elissa M Redmiles and Eszter Hargittai. 2018. New phone, who dis? Modeling millennials’ backup behavior. *ACM Transactions on the Web (TWEB)* 13, 1 (2018), 1–14. <https://doi.org/10.1145/3208105>.
- [26] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. 2016. How I Learned to be Secure: a Census-Representative Survey of Security Advice Sources and Behavior. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. Association for Computing Machinery, New York, NY, USA, 666–677. <https://doi.org/10.1145/2976749.2978307>
- [27] Elissa M Redmiles, Amelia Malone, and Michelle L Mazurek. 2016. I Think They’re Trying to Tell Me Something: Advice Sources and Selection for Digital Security. In *2016 IEEE Symposium on Security and Privacy*. IEEE, San Jose, USA, 272–288. <https://doi.org/10.1109/SP.2016.24>.
- [28] Elissa M. Redmiles, Noel Warford, Amritha Jayanti, Aravind Koneru, Sean Kross, Miraida Morales, Rock Stevens, and Michelle L. Mazurek. 2020. A Comprehensive Quality Evaluation of Security and Privacy Advice on the Web. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, 89–108.
- [29] Robert W. Reeder, Julia Ion, and Sunny Consolvo. 2017. 152 Simple Steps to Stay Safe Online: Security Advice for Non-Tech-Savvy Users. *IEEE Security and Privacy* 15, 5 (2017), 55–64. <https://doi.org/10.1109/MSP.2017.3681050>
- [30] Paul Ruggiero and Matthew A. Heckathorn. 2012. Data Backup Options. Accessed 6/9/24 [https://www.cisa.gov/sites/default/files/publications/data\\_backup\\_options.pdf](https://www.cisa.gov/sites/default/files/publications/data_backup_options.pdf).
- [31] Statista. 2022. Population of the United States in 2022, by educational attainment. <https://www.statista.com/statistics/240868/educational-attainment-in-the-us/> accessed in November 2024.
- [32] Statista. 2023. Verteilung der Bevölkerung in Deutschland nach beruflichem Bildungsabschluss im Jahr 2023. <https://de.statista.com/statistik/daten/studie/3276/umfrage/bevoelkerung-nach-beruflichem-bildungsabschluss/> accessed in November 2024.
- [33] Javier Suárez-Álvarez, Ignacio Pedrosa, Luis Lozano, Eduardo García-Cueto, Marcelino Cuesta, and José Muñiz. 2018. Using reversed items in Likert scales: A questionable practice. *Psicothema* 30 (05 2018), 149–158. <https://doi.org/10.7334/psicothema2018.33>
- [34] Jenny Tang, Eleanor Birrell, and Ada Lerner. 2022. Replication: How well do my results generalize now? The external validity of online privacy and security surveys. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX, Boston, USA, 367–385.
- [35] Tolga Tavlas. 2024. *CYBERSECURITY DICTIONARY for Everyone: 1250 Terms Explained in Simple English*. Independently published.
- [36] Rick Wash and Emilee Rader. 2015. Too Much Knowledge? Security Beliefs and Protective Behaviors Among United States Internet Users. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX, Ottawa, Canada, 309–325.
- [37] Gail H. Weems, Anthony J. Onwuegbuzie, and Daniel Lustig. 2003. Profiles of Respondents Who Respond Inconsistently to Positively- and Negatively-worded Items on Rating Scales. *Evaluation & Research in Education* 17, 1 (2003), 45–60. <https://doi.org/10.1080/14664200308668290>
- [38] Alma Whitten and J Doug Tygar. 1999. Why Johnny Can’t Encrypt: A Usability Evaluation of PGP 5.0.. In *USENIX Security Symposium*, Vol. 348. 169–184.
- [39] Yev. 2022. The 2022 Backup Survey: 54% Report Data Loss With Only 10% Backing Up Daily. <https://www.backblaze.com/blog/the-2022-backup-survey-54-report-data-loss-with-only-10-backing-up-daily/> Accessed 20/11/24.

## A CODES FOR DATA LOSS AND RECOVERY STORY

Table 11. Assigned codes for a data loss incident whose details the participants most easily recall, part 1 ( $N = 656$ )

Code	Description	<i>N</i>	%
<i>What was Lost?</i>			
All files	Everything was affected.	302	46.0%
Some files	Some data was affected, but it was not specified exactly which ones.	93	14.2%
Photos	Photos were affected.	117	17.8%
Documents	Documents were affected, e.g., a PDF file.	74	11.3%
Videos	Videos were affected.	19	2.9%
Messages	Chat histories or text messages were affected.	14	2.1%
Music	Music files were affected.	10	1.5%
Contacts	Contacts were affected.	6	0.9%
Mail	Mails were affected.	2	0.3%
Miscellaneous	Something was affected that does not fit into the other categories, e.g., game files	28	4.3%
<i>What Happened?</i>			
Broken device	Device broke, for example, when it was dropped, water spilled or the system crashed.	315	48.0%
Broken external hard drive	External hard drive that was used as additional memory for the device (storage space).	47	7.2%
Accidental deletion of some files	Some files were accidentally deleted.	43	6.6%
Everything accidentally deleted	All files were accidentally deleted.	7	1.1%
User error other	Another user error occurred, e.g., a file was not saved.	30	4.6%
Virus	A virus or malware got onto the device.	32	4.9%
Moving files to another device	Data was transferred, e.g., from the old device to another device.	21	3.2%
Moving files to clouds	Data was transferred, e.g., from the old device to the cloud.	4	0.6%
Lost device	Device was lost.	18	2.7%
Stolen device	Device was stolen.	17	2.6%
Software crashed	Software crashed and the file was not saved before.	13	2.0%
Update failure	An update failed and deleted the files.	7	1.1%
Cloud failure	Something broke in the cloud and data was deleted.	7	1.1%
Don't know	Unsure or cannot remember what exactly happened.	63	9.6%
Other incident	Something else has happened, e.g., a password was forgotten that was needed to access data, or a power failure caused current data not to be saved.	22	3.4%
<i>Barriers to Backing Up</i>			
No backup	There was no backup.	50	7.6%
Old backup	Backup was outdated.	27	4.1%
Broken backup	Backup was broken.	18	2.7%
Partial backup	Not all important data has been backed up. This code should be only assigned if the user has intentionally not saved all data beforehand.	9	1.4%
No money	Lack of money is mentioned why backup or recovery was not possible. For example, to pay for recovery software or services.	7	1.1%
No knowledge	The participant did not know at that time how to create a backup.	4	0.6%
Other backup problem	Another backup problem occurred.	1	0.2%

Table 12. Assigned codes for a data loss incident whose details the participants most easily recall, part 2 (N = 656)

Code	Description	N	%
<i>Recovery of Data</i>			
No recovery	Recovery was not possible.	64	9.8%
Partial recovery	Only some of the data could be recovered.	47	7.2%
Backup recovery	Data is recovered from a backup.	43	6.6%
Cloud recovery	Data is recovered from the cloud.	34	5.2%
Service Recovery	A service is hired to recover the data.	30	4.6%
Software recovery	Data is recovered with the help of dedicated software.	21	3.2%
Device recovery	Data is recovered from a different device, e.g., a smartphone.	11	1.7%
Update recovery	Data recovered after an update of the OS.	1	0.2%
Unclear recovery	It is not clear how exactly the data was recovered.	18	2.7%
Other recovery	Another method of recovery is used, e.g., a file that was previously deleted was restored from the recycle bin.	15	2.3%
<i>Other</i>			
Emotion	The loss is associated with strong emotions.	35	5.3%
Important files	It is highlighted that important or sensitive data was lost.	33	5.0%
System reset	Device's system was reset, deleting the data.	31	4.7%
Started doing backups	Started doing backup after a loss of data / device.	23	3.5%
Switched backup method	Backup method was changed.	2	0.3%
Multiple times	Data was lost multiple times from the same device.	7	1.1%
Other	Category in which all other comments are caught.	4	0.6%

## B EMOTIONAL IMPACT OF DATA LOSS AND RECOVERY EVENTS

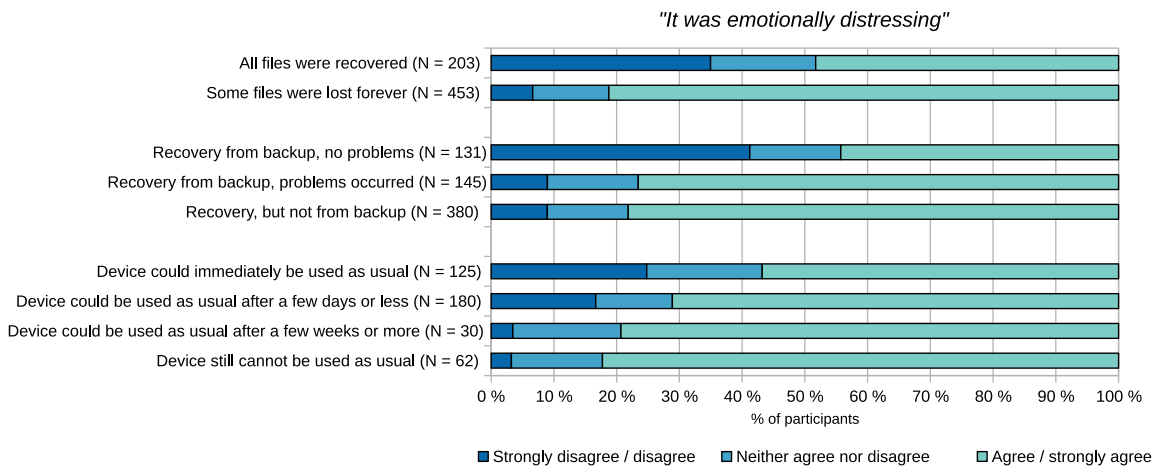


Fig. 12. Agreement with the statement "It was emotionally distressing" compared to the indication whether the data was restored from a backup, whether all data could be restored or not and how long it took until the device could be used again as usual (N = 656).

## C CODES FOR BACKUP CHARACTERISTICS

Table 13. Assigned codes of the qualitative results for backup characteristics that are important to the participants ( $N = 1423$ ).

Code	Description	<i>N</i>	%
Easy to use	The backup should be easy to create or the data should be easy to recover.	719	50.5%
Reliable	The backup should be able to reliably store and recover the data.	401	28.2%
Fast	The backup should be created quickly or the data should be restored quickly. It should be efficient.	327	23.0%
Secure	The backup should be safe from viruses or hackers.	272	19.1%
Cheap	The backup should be appropriate for the price, cheap or even free.	199	14.0%
Automation	The backup process should be performed automatically and regularly.	169	11.9%
Accessible	Access to the backup data should be easy. This code is also assigned when data availability is mentioned.	146	10.3%
Storage	The backup should have enough storage space.	95	6.7%
Complete	The backup should be complete and everything that was selected should be saved. This code is also assigned when all data should be saved.	55	3.9%
Compatible	The backup should be easily compatible with other devices. For example, the data from the backup should be easily recoverable on another device.	37	2.6%
No interruption	The backup process should not affect the use of the device.	32	2.2%
Notification	The backup should tell exactly what data is being backed up and when. It should communicate better what exactly it is doing.	28	2.0%
Control	It should be possible to control access to the backup so that no unauthorized persons can access the data.	27	1.9%
Options	Backup should offer more adjustable options so that backup and recover process can be modified.	26	1.6%
Up-to date	The backup should be up-to-date.	19	1.3%
Local	The backup should only be created locally and not be connected to the Internet.	15	1.1%
Many methods	A wide range of backup methods is provided.	13	0.9%
Other	A different characteristic is named that does not fit into the above categories.	47	3.3%
Don't know	The participant writes "Don't know" or something similar.	14	1.0%

## D CODES FOR WANTING AND NOT WANTING TO USE CLOUD BACKUP

Table 14. Assigned codes for the question if and why the participants want to backup all their data in the cloud ( $N = 1423$ ).

Code	Description	N	%
<i>Yes, I want to backup all my data in the cloud. (N = 486)</i>			
Backup	The cloud serves as another backup for the data. This code should be also assigned if fear of losing data is mentioned.	296	60.9%
Easy	Using the cloud is easy.	94	19.3%
Accessibility	All data is stored in one place and can be accessed from several places.	84	17.3%
New devices	The cloud can be accessed when setting up a new device.	30	6.2%
Different devices	The cloud can be accessed from different devices.	24	4.9%
More storage	The cloud has more storage space, so some data can be outsourced.	28	5.8%
Useful	The cloud is useful and helpful.	19	3.9%
No hardware	No additional hardware is needed.	3	0.6%
Yes-Other	Another positive aspect is mentioned.	28	5.8%
<i>It Depends. (N = 471)</i>			
Should be secure	All data in the cloud should be safe from loss and hackers.	79	16.8%
Appropriate cost	The cost of using cloud should be appropriate.	69	14.6%
Should be free	The cloud should be free to use.	37	7.9%
Should be easy	The usage of the cloud should be simple. Here it is perceived as too complicated.	40	8.5%
Depends on data	It depends on the specific data if the participant wants to upload it to the cloud.	32	6.8%
Should be accessible	Data in the cloud should always be accessible.	12	2.5%
Depends on device	It depends on the device that should be backed up. For example, some want to backup their phone only, but not their desktop computer.	9	1.9%
Customizable	The cloud should have adjustable options so that what data, how and where it is stored can be specified.	5	1.1%
Depends-Other	Another "it depends" aspect is mentioned.	34	7.2%
<i>No, I don't want to backup all my data in the cloud. (N = 466)</i>			
No trust	The cloud is perceived as too insecure considering storing data as they might disappear.	201	43.1%
Only important files	Only certain, mostly important, data should be stored - not all. This code should be also assigned if only some data should be stored.	182	39.1%
No need	There is no need to use the cloud for backup.	102	21.9%
Privacy concerns	The cloud is perceived as not trusted concerning private data. No particularly sensitive or private data should be put inside.	88	18.9%
Fear of hackers	This code is assigned when hackers or attackers are explicitly mentioned.	31	6.7%
No knowledge	The participants do not know how the cloud works or is used. This code is also assigned if the participant mentioned that the cloud is too complicated.	70	15.0%
Restricted space	Data would take up too much space in the cloud.	62	13.3%
No personal files	Very sensitive or personal files should not be stored.	60	12.9%
Multiple backup methods	It is mentioned that several backup methods are used.	44	9.4%
Cost too high	The costs for the cloud are considered to be too high.	38	8.2%
Performance concerns	Using the cloud is perceived as laborious and demanding on resources.	19	4.1%
Not always accessible	Participants fear losing access and control over all their data if they lose their cloud credentials.	7	1.5%
Internet needed	A good and stable Internet connection is required for the cloud, which is considered troublesome by the participants.	16	3.4%
No-Other	Another negative aspect is mentioned.	14	3.0%