"Now I'm a bit angry:" Individuals' Awareness, Perception, and Responses to Data Breaches that Affected Them

Peter Mayer*
peter.mayer@kit.edu
SECUSO / KASTEL
Karlsruhe Institute of Technology

Yixin Zou*
yixinz@umich.edu
University of Michigan

Florian Schaub
fschaub@umich.edu
University of Michigan

Adam J. Aviv
aaviv@gwu.edu
The George Washington University

Abstract

Despite the prevalence of data breaches, there is a limited understanding of individuals' awareness, perception, and responses to breaches that affect them. We provide novel insights into this topic through an online study (n=413) in which we presented participants with up to three data breaches that had exposed their email addresses and other personal information. Overall, 73% of participants were affected by at least one breach, 5.36 breaches on average. Many participants attributed the cause of being affected by a breach to their poor email and security practices; only 14% correctly attributed the cause to external factors such as breached organizations and hackers. Participants were unaware of 74% of displayed breaches and expressed various emotions when learning about them. While some reported intending to take action, most participants believed the breach would not impact them. Our findings underline the need for user-friendly tools to improve consumers' resilience against breaches and accountability for breached organizations to provide more proactive post-breach communications and mitigations.

1 Introduction

Data breaches, the disclosure of sensitive personal information to unauthorized parties, are on the rise [30, 63]. The average user has accounts with 191 online services [18]. Meanwhile, the Have I Been Pwned (HIBP) breach database lists over 480 breached online services and over 10M compromised accounts [29]. The Identity Theft Resource Center reported 1,108 breaches that occurred in the United States in 2019, which exposed over 164M sensitive records [30]. The sheer number of breaches makes it challenging to track the total number of records involved [35] and notify affected consumers [83]. Facing a plethora of data breaches [30, 63], consumers rarely take recommended protective measures in response [1, 31, 99].

Prior work has primarily studied consumers' general reactions to data breaches [1, 31, 37] or has focused on individual breaches in isolation such as the Equifax [99] and Target breaches [27, 41]. By contrast, we conducted an online study (n=413) in which we leveraged the HIBP database to present participants with, and have them reflect on, specific data breaches that had exposed their email address and other personal information. With this novel approach, we gathered 792 detailed breach-specific responses (up to three per participant), covering 189 unique breaches and 66 different exposed data types. Our quantitative and qualitative analyses contribute valuable insights into individuals' awareness, perception, and responses to specific data breaches that affected them. We further tease out interactions between individuals' awareness, concern, and self-reported action. Our findings answer the following research questions:

RQ1 [Breach status] *What factors influence the likelihood that an email address is involved in a data breach?*

Overall, 73% of our participants experienced at least one breach and 5.36 breaches on average. An email address's likelihood of being exposed in a breach significantly correlated with the email account's age and utilization.

RQ2 [Perception] *What do participants perceive as the causes of being involved in data breaches and related impacts, and to what extent do their perceptions align with reality?*

Only 14% of our participants accurately attributed the cause of being affected by a breach to external factors such as breached organizations and hackers. Others blamed their email or security behaviors for making themselves a victim or viewed breaches as inevitable. Most participants expected little impact from shown breaches despite realizing certain risks.

RQ3 [Awareness] *What factors influence participants' awareness of data breaches that affected them?*

Participants were unaware of most data breaches presented (74%). Those who knew they were affected by a specific breach had primarily learned about it from the breached

*Peter Mayer and Yixin Zou contributed equally to this research.

organization or third-party services. Participants were more likely to be aware of older rather than recent breaches.

RQ4 [Emotional response] *What are participants' emotional responses to data breaches that affected them?*

Most participants rated their concern regarding breaches as low (56% slightly/somewhat concerned, 19% no concern). Certain breached data types such as physical address and password raised more concern than others. Participants expressed emotions ranging from upset, angry, annoyed, frustrated, surprised (or not) to violated and fatigued.

RQ5 [Behavioral response] *What factors influence participants' likelihood to take action in response to data breaches that affected them?*

Participants reported having already or being very likely to change their passwords and review credit reports/financial statements in response to over 50% of shown breaches. Participants were more likely to take action with increased concern and prior awareness, suggesting that better communication about breaches could increase individuals' tendency to take protective actions.

Our findings demonstrate the need for more proactive communications of data breaches and stronger protections for affected individuals. Rather than burdening consumers to take action, breached organizations should be held responsible for increasing awareness and providing appropriate mitigations. Furthermore, our findings highlight the need for usable privacy tools to help affected individuals be more resilient against future breaches.

2 Background and Related Work

Data breaches. Data breaches have multifaceted consequences. Breached organizations can bear substantial costs to repair the aftermath, including patching system vulnerabilities, compensations to affected individuals, and resolving potential lawsuits [71, 72]. There are also invisible and hard-to-measure costs in rebuilding the breached organization's reputation [39, 94] and affected individuals' trust [1, 12, 49]. For affected individuals, exposed data puts them at risk of account compromise [18, 66, 77, 87], phishing [59], and identity theft [70, 74, 81]. Though it may take years before leaked data is misused, the harm can be profound when it happens. For instance, victims of identity theft may have ruined credit reports or have to file for bankruptcy due to abuse of credit [5]. Identity theft is also traumatizing: in a 2017 survey by the Identity Theft Resource Center [43], 77% of respondents reported increased stress levels, and 55% reported increased fatigue or decreased energy. Thus, some researchers [16, 81] have argued that data breaches cause compensable harms due to the substantial risk of future financial injury and the emotional distress imposed on victims.

Breached organizations are often legally required to notify affected victims [22, 61] and offer compensations such as discounts [13] or free credit/identity monitoring [76]. Services like HIBP [29] and Firefox Monitor [53] examine third-party breach reports and notify signed-up users. Some companies automatically reset passwords for users whose credentials appeared in password dumps [26, 95]. Additional measures for victims include two-factor authentication (2FA) that increases the difficulty of misusing leaked credentials and warnings that flag social engineering and phishing attacks [46, 60]. Nevertheless, no solution is perfect: attackers can bypass 2FA without obtaining the secondary token [19, 32], and phishing warnings have low adherence rates [3, 4, 21].

Security mental models and behaviors. How individuals perceive the causes and impacts of data breaches relates to mental models of security and privacy. Mental models — an individual's internalized representation of how a system works [56] — have been studied for computer security [91], security warnings [9], smart home security [97], and the Internet [36]. Respective studies consistently find that unawareness and misconceptions of security risks create hurdles for adopting effective mitigation strategies. Even when individuals correctly assess risks, they may still not react accordingly due to bounded rationality and cognitive biases [2] or not having experienced negative consequences [100].

We investigate two aspects that may impact how individuals respond to data breaches: *awareness*, i.e., whether and how individuals learn about a breach, and *perception* regarding a breach's potential causes and impacts. For awareness, prior research has documented various channels individuals leverage to learn about security advice, including media, peers, family, workplace, and service providers [15, 65, 67]. For data breaches specifically, respondents of RAND's 2016 US national survey [1] reported first learning of a breach from the breached organization's notification (56%), media reports (28%), or third-parties (16%). Additionally, prior research has shown that consumers understand the potential impacts of data breaches, such as identity theft and personal information leakage [31, 37, 99]. Our study complements these findings by prompting participants to reflect on both causes and impacts of specific breaches that affected them, providing insights on how these perceptions link to their emotions and behaviors.

Consumer reactions to data breaches. Data breach victims are advised to take a range of actions depending on the information exposed [85, 86, 90], such as changing passwords if account credentials are exposed or requesting new cards and reviewing statements if financial information is exposed. In the US, victims are further urged to place a credit freeze, check credit reports, and file taxes early if their Social Security number (SSN) is exposed [47, 84, 85].

Nevertheless, studies on breaches in general [1, 31, 37] and on specific breaches [27, 41, 88, 99] show that con-

sumers rarely take recommended protective measures in response [31, 99, 100]. While consumers report increased concern about identity theft [6, 31] and diminished trust in the breached organization [12, 55], such risk perception and attitudinal change often do not result in action. Consumers tend to accept compensations provided by the breached organization [1, 51] but do not go further; they continue using existing credit cards [51] and the same password for different accounts [25], thereby fueling credential stuffing attacks that cause account compromises [30].

Several studies have examined the determinants of consumers' behavioral reactions to data breaches: knowledge of available measures [99], perception of clear evidence indicating being affected [50], cognitive biases [99], peer influence [14, 41], and media coverage [15]. Tech-savvy and non-tech-savvy individuals also differ in their needs for guidance related to mitigating actions [6]. Furthermore, breach notifications to victims are often ambiguous in communicating risks and priority among recommended actions [8, 89, 98]. These issues, coupled with the overwhelming amount of security advice for end-users [68, 69], may pose challenges for affected individuals to act on provided advice.

Methodologically, prior work primarily asked participants to recall past experiences with generic breaches [1, 31] or describe intended reactions in hypothetical scenarios [28,37]. By contrast, we apply a novel approach to examine participants' responses to specific breaches that exposed their information. Our study covers a multitude of breaches varying in size and types of exposed information rather than one breach as a case study [27, 51, 88, 99]. Our approach increases ecological validity and mitigates recall bias as participants are confronted with breaches that affect them. Similar reflection studies have yielded insights into users' attitudes and behaviors in other contexts, such as password creation behaviors [58, 92] and reactions to online tracking [93] or advertising inference [64].

3 Method

Our study addresses our five research questions as follows. To identify what factors influence an email address's likelihood of being involved in a breach (RQ1), we collected details about participants' email usage and demographics. To identify perceptions regarding the causes of being involved in a breach and related consequences (RQ2), we asked participants to speculate why their email address may have or have not been involved in any data breaches, and any associated impacts they expect or have experienced. For each specific breach, we asked participants if they were previously aware of it and, if so, how (RQ3). To assess emotional responses, we asked participants to describe how they feel about the breach and rate their concern (RQ4). We further asked participants to self-report what they did in response to the breach and rate the likelihood of taking (or having taken) ten provided actions (RQ5). We ran regression models to examine the relationship between email usage, breached data types, awareness, concern, and behavioral reactions. Our study was approved by our Institutional Review Boards (IRB).

3.1 Survey Instrument

As we were motivated to understand participants' responses to real-world breaches at scale, we conducted an online survey with data pulled from Have I Been Pwned (HIBP).[1] We built a survey platform which queried the HIBP web service API using email addresses provided by study participants. To protect participants' confidentiality, we only maintained email addresses in ephemeral memory to query HIBP. At no point did we store participants' email addresses. We then used the query results, i.e., the breaches in which a participant's email address was exposed, to drive the remainder of the survey. The survey consisted of three main parts (see Appendix A).

Part 1: Email address-related questions. After consenting, we asked participants for their most commonly used email address. We clearly noted that the email address will only be used to query HIBP and that we will never see it (Appendix A.2). Once a participant entered an email address, we asked a few questions about it. Participants who indicated that the email address belonged to someone else or was fabricated were given the option to enter a different email address or leave the study. Next, we asked participants about their email habits as a potential influencing factor of the email's involvement in breaches (RQ1). This included frequency of checking their email, primary use of the account (professional/personal correspondence or account creation), how long it has been used, and the number of other email accounts the participant used. We then used the provided email address to query HIBP.

Part 2: Breach-related questions. We next informed participants whether their email address was exposed in any data breaches without stating the specific number or giving more details. To answer RQ2, we asked participants to speculate why their email address was or was not part of data breaches. Participants whose email address was not part of any breach were given the opportunity to enter a different email address until a provided email address had associated breaches. If they did not provide another email, they continued with part 3.

We randomly selected up to three breaches, displayed one by one, to ask breach-related questions while limiting potential fatigue. We displayed a breach's description, logo, name, and types of compromised data as provided by HIBP (Figure 1). We explicitly stated that these were actual breaches (see Appendix A), and no participants doubted the validity of shown breaches in their qualitative responses. For each breach, we asked about participants' awareness (RQ3), emotional response (RQ4), and actions taken or intended to take (RQ5).

[1] https://haveibeenpwned.com

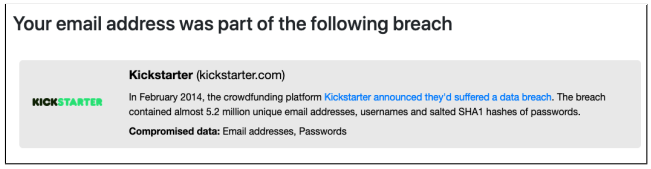

Figure 1: Sample breach information shown to participants.

For emotional response, participants provided open-ended responses, then rated their concern level on a 5-point Likert scale regarding the breach in general and for each type of exposed data. For behavioral response, participants described their reactions (open-ended) before rating their intention to take (or whether they had taken) ten provided actions sourced from prior work [85, 86, 90]. The respective breach information was visible at the top of the page when participants answered all these questions.

Part 3: Demographics, attention check, and debrief. We collected participants' demographics including age, gender, education, whether they had a background in IT or law, and household income. We also included two attention check questions: one asking them to identify the name of a breach shown during the study (only for participants whose email address was part of at least one breach), and a generic attention check (see Appendix A.4). Finally, we showed participants a list of all breaches associated with their provided email address and links to resources on data breach recovery to help them process and act on this potentially new information.

3.2 Recruitment

We recruited participants via Prolific,[2] an online research platform similar to Amazon Mechanical Turk with more demographically diverse subjects [57], between August and October 2020. We balanced participants' age and gender distributions in data collection. After the first 171 participants, we realized and corrected a storage error that caused missing data in income and ratings for taken/intended actions. We note in Section 5 how we accounted for this in our analyses. Participants were compensated $2.50 for an average completion time of 13.37 minutes ($11.22/hour).

3.3 Analyses

We collected data from 416 participants; three participants were excluded as they did not respond to any open-ended questions meaningfully, resulting in 413 participants in total. We based our sample size on our planned analyses: Bujang et al. [11] suggest n=500 or n=100+50×#IVs as the minimum sample size for logistic regressions. For the linear regression (RQ4), G*Power suggests n=127 for detecting medium effects (f^2=.15), with α=.05, β=.80. With 413 participants (435 email-specific responses; 792 breach-specific responses) we met or exceeded these thresholds.

97% of participants passed our generic attention check. Of the 302 participants who were shown at least one breach, only 55% passed the breach-specific attention check, whereas the rest chose "none of these" (42%) or a decoy option (3%). We reviewed open-ended responses from participants who failed this attention check, and all of them were detailed and insightful. We also did not find significant correlations between this attention check's performance and participants' breach-specific responses about awareness (chi-squared test, $\chi(1)$=.06, p=0.8), concern level (Mann Whitney test, W=58395, p=0.2), and whether they had taken action (chi-squared test, $\chi(1)$=.29, p=0.6). Thus, we did not exclude any of these participants as our findings suggest the question was not a reliable exclusion criterion.

Qualitative analysis. We analyzed participants' open-ended responses using inductive coding [75]. For Questions 7, 10, 14, 16, and 18, a primary coder created an initial codebook based on all responses. Multiple coders then iteratively improved the codebook. A second coder analyzed 20% of responses to each question to ensure high inter-rater reliability [45]. Cohen's κ were 0.89 (Q7), 0.73 (Q10), 0.74 (Q14), 0.81 (Q16), and 0.78 (Q18). We resolved all coding discrepancies through discussions. Appendix B includes the codebook, with common themes highlighted.

Statistical analysis. We conducted regressions to identify influential factors with respect to breach status (RQ1), awareness (RQ3), emotional response (RQ4) and behavioral response (RQ5). We included a random-intercept for individual participants to account for repeated observations between multiple breaches. However, for models corresponding to RQ1 the random effects were close to zero and caused a boundary singularity fit, so we conducted single-level regressions instead. For all models, we treated participant demographics (age, gender, education, occupational background) as control variables: we report a model's output with participant demographics when it has a significantly better fit than the model without; otherwise, we opt for the simpler model in reporting the results. We treated participants' responses of concern level on a 5-point Likert scale as a continuous variable in our regressions, which has limitations, as we discuss below.

3.4 Limitations

As with most surveys, parts of our findings rely on self-reported data, which is prone to biases. For instance, prior work has shown a gap between self-reported behavioral intentions and actual behaviors in security contexts [34] and beyond [78]. We do not imply that all participants would take actions they reported. Nevertheless, participants' self-reported

[2] https://prolific.co

intentions to act can inform future research and mechanism design to better protect consumers against data breaches.

HIBP's API does not return breaches marked sensitive such as those involving adult sites. Accessing these breaches requires sending a confirmation message to participant-provided email addresses for ownership verification. We decided not to do this as it may suggest to participants that we store their email addresses even though we do not.

Our study only included data breaches involving email addresses, which may not represent all breaches (e.g., only 4% of breaches recorded by Privacy Rights Clearinghouse [63] included email addresses). Relatedly, the email-focused nature of these breaches means it is difficult to track whether and how breached organizations in our sample notified affected individuals and how that impacts consumer reactions, because existing breach notification databases mostly document letter-based notifications [98]. Future research can look into breaches that expose a broader range of data types and consider organizations' handling of breaches when feasible.

Regarding our analyses, we considered several options of treating the Likert responses of concern level: ordinal, nominal, or continuous. Treating concern as ordinal would introduce square and cubit effects into the model — these effects are difficult to interpret and inconsistent with the scale. Treating concern as nominal would lose information about the scale's ordering and prevent comparisons across all levels (e.g., with "not at all concerned" as the baseline, the regression would not describe the difference when moving up or down the scale between "slightly concerned" and "extremely concerned"). Treating concern as continuous would require a more cautious interpretation of the p-values in the analysis, and it assumes equal differences between the scale items. After discussions with our university's statistical consulting service, we followed their advice and decided to treat concern as a continuous variable. While this comes with the limitations mentioned above, it also allows a more straightforward and meaningful interpretation of results, which we prioritize to make the results more accessible.

4 Data Description

Participant profile. Table 1 summarizes our 413 participants' demographics and breach status. Our participants were almost evenly distributed between men and women but skewed educated and younger. 122 (30%) described having a background in information technology; 25 (6%) in law.

In total, participants provided 435 email addresses. 421 (97%) accounts were solely owned by the participant, and ten were shared with someone else. Four were either someone else's account or a made-up address for the study, and so were removed from the data. Participants whose initial email address was not exposed in any breach could scan another: 393 participants (95%) scanned only one email address, 18 scanned two addresses, and only two scanned three addresses.

	Total	Num. (%) W/ Breaches	Num. (%) W/o Breaches	Avg. (Med./Std.) Breaches
Men	**199**	139 (70%)	60 (30%)	4.49 (2/5.97)
Women	**212**	162 (76%)	50 (24%)	6.11 (4/6.28)
Non-Binary	**2**	1 (50%)	1 (50%)	11.00 (11/11.00)
18-24	**77**	56 (73%)	21 (27%)	3.90 (2/5.15)
25-29	**51**	35 (69%)	16 (31%)	4.25 (2/4.90)
30-34	**42**	33 (79%)	9 (21%)	6.55 (3/8.72)
35-39	**49**	29 (59%)	20 (41%)	4.63 (1/7.05)
40-44	**45**	26 (58%)	19 (42%)	4.36 (2/5.04)
45-49	**32**	29 (91%)	3 (9%)	6.59 (4/6.05)
50-54	**39**	30 (77%)	9 (23%)	6.72 (6/6.16)
54-59	**34**	30 (88%)	4 (12%)	6.12 (5/4.82)
60-64	**27**	19 (70%)	8 (30%)	6.52 (3/6.85)
65+	**17**	15 (88%)	2 (12%)	8.24 (8/6.06)
Some High School	**1**	0 (0%)	1 (100%)	0.00 (0/0.00)
High School or Equiv.	**46**	35 (76%)	11 (24%)	4.59 (3/4.61)
Some College	**88**	70 (80%)	18 (20%)	5.67 (3/6.63)
Associate (voc./occ.)	**14**	14 (100%)	0 (0%)	8.07 (6/6.51)
Associate (aca.)	**20**	19 (95%)	1 (5%)	6.10 (4/5.99)
Bachalor	**140**	108 (77%)	32 (23%)	6.04 (4/6.56)
Masters	**83**	46 (55%)	37 (45%)	4.10 (2/5.68)
Professional	**5**	4 (80%)	1 (20%)	11.60 (13/7.71)
Doctorate	**16**	6 (38%)	10 (62%)	1.44 (0/2.26)
IT Background	**122**	67 (55%)	55 (45%)	3.82 (1/6.30)
No IT Background	**278**	224 (81%)	54 (19%)	5.91 (4/6.06)
Prefer not to say	**13**	11 (85%)	2 (15%)	8.00 (9/6.41)
Law Background	**25**	14 (56%)	11 (44%)	5.80 (2/9.63)
No Law Background	**374**	278 (74%)	96 (26%)	5.29 (3/5.93)
Prefer not to say	**14**	10 (71%)	4 (29%)	6.36 (5/6.25)
No Data	**170**	115 (68%)	55 (32%)	4.45 (2/6.21)
<$15K	**16**	15 (94%)	1 (6%)	7.81 (4/8.59)
$15K-$25K	**22**	20 (91%)	2 (9%)	6.77 (4/5.79)
$25K-$35K	**28**	26 (93%)	2 (7%)	5.89 (3/5.37)
$35K-$50K	**26**	19 (73%)	7 (27%)	4.58 (2/5.35)
$50K-$75K	**45**	40 (89%)	5 (11%)	8.04 (7/6.50)
$75K-$100K	**38**	28 (74%)	10 (26%)	6.95 (4/6.61)
$100K-$150K	**37**	22 (59%)	15 (41%)	4.05 (2/4.63)
>$150K	**24**	13 (54%)	11 (46%)	3.92 (2/5.34)
Total	**413**	**302 (73%)**	**111 (27%)**	**5.36 (3/6.23)**

Table 1: Participant demographics and breach status (n=413).

For the 431 owned or shared email accounts, we further asked participants how long they had been using the email account, how frequently they checked it, and what they primarily used it for. The majority of email accounts were used for an extended period (mean: 8.75 years, median: 8). Most (81%) were checked daily; the rest were checked less frequently (14% weekly, 4% monthly, and 1% yearly). Participants reported multiple uses for their email address (mean: 2.74, median: 3): 74% were used for personal correspondence, followed by signing up for medium-sensitive accounts like social media (68%), signing up for sensitive accounts like banking (51%), signing up for low-value accounts (49%), and professional correspondence (32%).

Overview of breaches. We observed 189 unique breaches across 431 email addresses queried against HIBP. 302 (70%) email addresses, or 73% of participants, were exposed in one or more breaches. The average number of breaches per email address was 5.12 (median: 3, sd: 6.21, max: 46), or 5.36 per participant (median: 3, sd: 6.23). The number of breaches per email address formed a long-tail distribution: 34% of email addresses appeared in 1 to 5 breaches, and only 2% were

associated with 21 or more breaches.

For the 189 unique breaches, we examined their date, the total amount of breached accounts, and the types of compromised data according to HIBP. The majority (69%) of breaches occurred in 2015–2019; 15 breaches occurred in 2020. The average number of breached accounts captured by HIBP was 46.52M (median: 4.79M; sd: 125M), indicating a distribution skewed by several large breaches (max: 772.90M). 66 different data types were leaked in our sample's breaches. The average number of leaked data types per breach was 4.86, and the maximum was 20 (median: 4, sd: 2.58). Aside from participants' email addresses (which were present in all breaches as HIBP uses them as references), the other commonly breached data types included passwords (162, 86%), usernames (110, 58%), IP addresses (82, 43%), names (74, 39%), and dates of birth (47, 25%). The frequency distribution of data types in our sample's breaches falls off steeply (see Figure 2), suggesting a broad range of leaked data types with a much smaller set of commonly leaked data.

We used Cisco's website content taxonomy[3] for cross-referencing breached organizations' industry, excluding 25 (13%) non-applicable cases.[4] Gaming companies were represented the most in our sample (40, 21%). Other represented industries included general business (17, 9%), computers/Internet (16, 8%), shopping (10, 5%), and online communities (10, 5%). We used Alexa's ranking of global websites[5] as of October 14, 2020 as a proxy for a breached organization's popularity.[6] Excluding 33 organizations with missing data, the average ranking was 650.73K (median: 24.85K, sd: 1,768K). 19 organizations appeared in the top 1K list, indicating that while the majority of organizations in our sample were not mainstream, a few were relatively well-known.

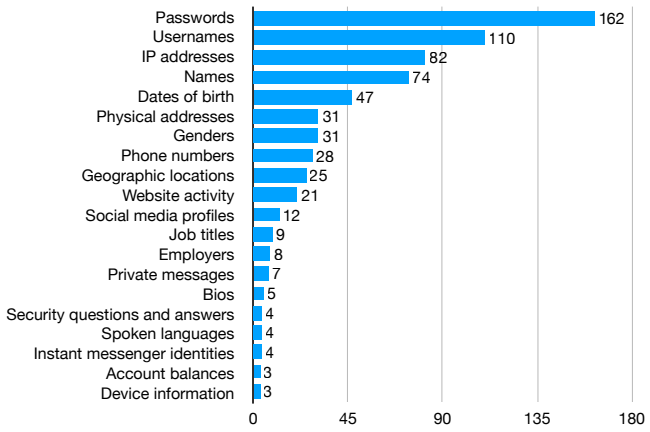

Figure 2: Frequency of the leaked data types for 189 breaches, excluding email address (appears in all breaches). 44 other types occurring twice or fewer.

5 Results

5.1 RQ1: Likelihood of Breaches

We conducted a logistic regression on whether an email address had been breached in relation to the email account's age, checking frequency, and purpose of use. Results in Table 2 show that an email address was significantly more likely to be breached as the account's age in years increased (OR_{age}=1.35, p<.001), as it was checked daily instead of weekly (OR^{weekly}_{daily}=2.30, p=.03), and as it was used for personal correspondence (OR^{no}_{yes}=2.13, p=.02). Additionally, the significant intercept indicates that an email address was significantly unlikely to be associated with any breach if the email account was just created, checked weekly, and not used for any correspondence or account creation purposes ($OR_{intercept}$=0.14, p=.002). Essentially, the less frequently used and newer an email address is, the less likely it is to be exposed in a data breach.

We further conducted a quasi-Poisson regression on the number of breaches per email address with the same independent variables as above. We chose quasi-Poisson regression because the dependent variable is count data with a skewed distribution [96]. Results in Table 3 show how the number of breaches increases with an email account's age: for every one year of increase in age, the expected number of breaches increases by a factor of $exp(0.08) = 1.08$ (p<.001). In other words, the number of breaches increases 8% per-year of use, compounding yearly (see Figure 3). A possible explanation is that the older an email address is, the more it has been used for account registrations, which increases its presence in organizations' databases. The significant intercept in Table 3 confirms this finding: a new and rarely used email address is more immune to breaches. Furthermore, the number of breaches per email address differed among age groups: compared to young adults (18-34), the number of breaches decreases by a factor of $exp(-0.29) = 0.75$ (p=.045) for middle-aged adults (35-54) and by a factor of $exp(-0.35) = 0.71$ (p=.02) for older adults (55+).

RQ1: What factors influence the likelihood that an email address is involved in a data breach? Our results suggest that an email account's age, checking frequency, and purpose of use are significant factors correlating with the email address's presence in a breach. Both models capture email age's influences: for each year of increase, the email address is

[3] https://talosintelligence.com/categories

[4] These breaches were spam lists or aggregate credential stuffing lists, or the breached organizations were no longer active.

[5] https://alexa.com/topsites

[6] We used rankings at the time of analysis rather than historic ranking (i.e., the ranking when the breach occurred) because (1) Alexa only provides ranking data for the last four years; and (2) we anticipate that current ranking would better reflect participants' impression of the organization's popularity at the time when they took our study.

Table 2: Logistic regression for breach status of an email address (leaked vs. not leaked).

	Est.	OR	95% CI	p-value
(Intercept)	−1.95	0.14	[0.04, 0.49]	.002
Freq. Checked daily (vs. weekly)	0.83	2.30	[1.07, 4.99]	.03
Prof. Corr. yes (vs. no)	−0.02	0.98	[0.51, 1.87]	.94
Pers. Corr. yes (vs. no)	0.76	2.13	[1.13, 4.03]	.02
Acct. Creat. yes (vs. no)	0.31	1.36	[0.60, 3.07]	.46
Email age years	0.30	1.35	[1.26, 1.46]	$< .001$
Age: 35-54 (vs. 18-34)	−0.51	0.60	[0.29, 1.23]	.16
Age: 55+ (vs. 18-34)	−0.60	0.55	[0.27, 1.10]	.09
Gender: men (vs. women)	−0.24	0.79	[0.43, 1.45]	0.45
Edu.: =Bach. (vs. <Bach.)	0.25	1.28	[0.65, 2.53]	0.48
Edu.: >Bach. (vs. <Bach.)	−0.62	0.54	[0.25, 1.16]	.11
Occu.: IT/law yes (vs. no)	−0.51	0.60	[0.31, 1.17]	.14

Table 3: Quasi-poisson regression regarding the number of breaches per email address.

	Est.	Exp (Est.)	SE	p-value
(Intercept)	0.67	1.94	0.26	.01
Freq. Checked daily (vs. weekly)	0.36	1.43	0.19	.06
Prof. Corr. yes (vs. no)	−0.11	0.89	0.12	.33
Pers. Corr. yes (vs. no)	0.29	1.34	0.15	.06
Acct. Creat. yes (vs. no)	−0.18	0.83	0.15	.22
Email age years	0.08	1.08	0.01	$< .001$
Age: 35-54 (vs. 18-34)	−0.29	0.75	0.14	.045
Age: 55+ (vs. 18-34)	−0.35	0.71	0.14	.02
Gender: men (vs. women)	−0.18	0.84	0.12	.13
Edu.: =Bach. (vs. <Bach.)	0.17	1.18	0.12	.18
Edu.: >Bach. (vs. <Bach.)	−0.17	0.84	0.16	.29
Occu.: IT/law yes (vs. no)	−0.05	0.95	0.14	.70

1.35x more likely to be part of a breach or gains 1.08x more breaches than the previous year. Conversely, the significant intercept in both models suggests that a new and rarely used email address is less likely to be involved in a breach. While these results are somewhat intuitive, they indicate the pervasiveness of data breaches: most email addresses queried in our study had appeared in one or more breaches even though they were only used in ordinary ways.

5.2 RQ2: Perceived Causes and Impacts of Being Affected by Breaches

We asked participants to speculate why or why not their email address was part of a data breach and name any experienced impacts or anticipated future impacts from a specific breach.

Perceived reasons for being affected by breaches. We analyzed 302 open-ended responses to Question 10 in which participants speculated why their email address was exposed in one or more data breaches. The most common explanation, cited in 159 (53%) cases, was that it was due to participants' own email-related practices. Specifically, 70 (23%) mentioned using the email address to sign up for many different sites (e.g., *"it's on the website of every business I have an online relationship with"*). Another 31 (10%) mentioned the email's age as a relevant factor, saying it had been used for a long time. 23 (8%) expressed that breaches were inevitable, especially for an old or widely-used email address (e.g., *"there are a lot of companies or organizations that have my email [address] and chances are one of them is going to get hacked"*). Furthermore, in 31 (10%) cases, participants mentioned using the email to sign up for seemingly sketchy websites, sometimes with a clear intention to do so despite knowing that the website might be insecure.

Participants mentioned other insecure behaviors as potential reasons for being affected by a breach in 31 (10%) cases. 13 cases referred to password-related behaviors, such as using simple passwords, reusing a password across accounts, or not changing passwords frequently. Incautious clicking behavior was mentioned five times (e.g., *"because I was not careful with what emails I clicked"*). Other participants indicated their exposure to breaches was due to infrequent monitoring of the email account, easily guessed answers for security questions, or being signed into the email account for too long. While these are indeed insecure behaviors, password choices do not impact one's likelihood of being involved in a breach; they impact a breach's consequences by increasing the possibility of account hijacking due to credential stuffing. Similarly, clicking on untrustworthy links may make the email address appear in spam lists, which will be reported by HIBP if found on the public web. However, this action on its own does not

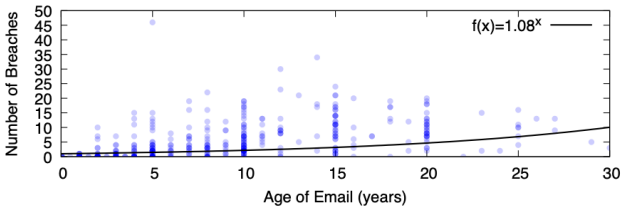

Figure 3: Number of breaches vs. age of email address (years); curve represents an 8% increase in number of breaches per year as estimated by the quasi-Poisson regression.

increase one's vulnerability to breaches.

Only 42 (14%) of participants accurately attributed the cause of being affected by a breach to external factors unrelated to their behaviors. 26 (9%) blamed it on lax security measures by the breached organization (e.g., *"these companies did not try hard enough to keep information private"*). 16 (5%) blamed it on bad actors such as hackers and scammers targeting the breached organization (e.g., *"hackers are devious devils and learn to adapt faster than organizations can protect users"*). Another 15 (5%) suspected their email address was sold by the breached organization or a third party. Nevertheless, nine participants incorrectly placed blame on their email provider's security (e.g., *"I feel like Hotmail has poor security and cannot block as many spam emails compared to Gmail"*).

Perceived reasons for not being affected by breaches. Question 7 asked participants to speculate why their email address was *not* involved in any data breach. Among the 136 provided responses, 78 (57%) mentioned cautious email practices. Specifically, 31 (23%) reported using their email address to sign up for trusted sites only, sometimes with careful examination of the website (e.g., *"I try as much as possible to scrutinize websites before dropping any of my details"*). 18 (13%) mentioned that their email address was relatively new or did not get used much, which is indeed a relevant factor, as shown by our regression results in Section 5.1. Ten further mentioned limiting the email to specific purposes, such as correspondence with friends and family members only.

Eight participants described using multiple email accounts for different purposes, e.g., using one email address for correspondence exclusively and another for account registration on "low-value" sites. Such behavior would likely reduce the likelihood of breaches involving high-value email addresses. However, breaches involving low-value email addresses may still have real impacts such as account hijacking.

21 (15%) participants cited their security practices as reasons for not being affected. Nine participants mentioned their password practices, such as using strong/unique passwords and changing passwords regularly. Less frequently mentioned were two-factor authentication, anti-virus, firewall, and VPN. None of these behaviors are likely to prevent data breaches despite potentially having other positive security outcomes.

Experienced and anticipated impacts of data breaches. Participants with at least one breach were asked to describe a given breach's experienced or potential impacts (Question 16). Of the 792 responses, more than half assessed the breach's impact as none (343, 43%) or very little (85, 11%); another 77 (10%) were unsure. Only 19 (4%) breaches were perceived as having a large impact. In 135 (17%) cases, participants described emotional feelings without naming concrete impacts, such as *"no impact just rage."*

In 149 (19%) instances, participants described specific experienced impacts or anticipated future impacts. The most prevalent was an increase in spam emails, text messages, etc. Some participants reported scam phone calls, and others anticipated identity theft as a potential impact (e.g., *"I suppose now that someone has all that information about me they could impersonate me, open credit lines in my name, scam my family and friends"*). Participants who had experienced adverse events described emotional stress and resulting behavioral changes, such as avoiding phone calls due to frequent scams or frequently checking emails for suspicious activities after account compromises.

Notably, participants with and without experienced impacts differed in assessing the impact's severity. Most participants who described anticipated impacts but had not experienced them did not foresee real consequences (e.g., *"the only things that [would] really happen is . . . scammers . . . occasionally attempt to access some of my older accounts that hold no sensitive information"*). This underlines that participants' perception of impacts after being affected by breaches largely depends on individual circumstances. The finding also aligns with prior work [99, 100] showing that people don't adopt secure behaviors until experiencing actual harms.

RQ2: What do participants perceive as the causes of being involved in data breaches and related impacts, and to what extent do their perceptions align with reality? Our results indicate that relatively few participants (42 out of 302, 14%) correctly attributed the cause of their victimhood to external factors such as the breached organization and hackers. Instead, most participants referred to their insecure behaviors related to email, passwords, etc., in explaining why their email address appeared in a breach. Most participants reported little to no experienced or anticipated impacts. When participants named concrete consequences, they mostly referred to spam and identity theft, though the perceived severity varied substantially.

5.3 RQ3: Awareness of Breaches

Among the 792 breach-specific responses, 590 (74%) reported unawareness of being affected by the breach before our study. Only 143 (18%) reported prior awareness, and

Table 4: Logistic regression regarding prior breach awareness.

	Est.	OR	95% CI	p-value
(Intercept)	−4.24	0.01	[0.002, 0.09]	$< .001$
Freq. Checked daily (vs. weekly)	0.31	1.37	[0.45, 4.16]	.58
Prof. Corr. yes (vs. no)	−0.06	0.94	[0.45, 1.98]	.88
Pers. Corr. yes (vs. no)	0.22	1.25	[0.50, 3.10]	.63
Acct. Creat. yes (vs. no)	0.77	2.15	[0.70, 6.63]	.18
Email age years	0.04	1.04	[0.98, 1.11]	.17
Breach age years	0.20	1.22	[1.09, 1.35]	$< .001$
Age: 35-54 (vs. 18-34)	−0.41	0.66	[0.27, 1.61]	.36
Age: 55+ (vs. 18-34)	−0.94	0.39	[0.15, 1.00]	.049
Gender: men (vs. women)	0.74	2.09	[1.00, 4.37]	.049
Edu.: =Bach. (vs. <Bach.)	−0.79	0.45	[0.20, 1.00]	.051
Edu.: >Bach. (vs. <Bach.)	−0.18	0.84	[0.31, 2.22]	.72
Occu.: IT/law yes (vs. no)	0.50	1.65	[0.72, 3.77]	.23

the other 8% were unsure. Participants who were previously aware of the breach mostly learned about it from the breached organization (45, 31%) or third-party notification services (45, 31%). Less common sources included news media (17, 12%), credit/identity monitoring services (14, 10%), bank or credit card companies (3, 2%), experiencing adverse events (3, 2%), and someone else (3, 2%). In nine instances, participants could not remember how they learned about the breach.

Using a mixed-effect logistic regression to identify factors that might impact awareness (excluding "unsure" responses), we included the same email-related factors from Table 2 as independent variables. Additionally, we included breach age (i.e., the time lapse between a breach's occurrence and the participant taking our study), hypothesizing that participants are more likely to recall and report awareness of recent breaches.

Results in Table 4 show a significant intercept, indicating that participants were more likely to be unaware of a breach if they have a newer email address and the breach just occurred ($OR_{intercept}$=0.01, $p<.001$). Participants were also significantly more likely to be aware of a breach as the breach's age in years increased (OR_{breach_age}=1.22, $p<.001$). Older participants were less likely to be aware of breaches than young participants (OR_{55+}^{18-34}=0.39, p=.049), and men were more likely to be aware of a breach than women in our sample (OR_{men}^{women}=2.09, p=.049), though p-values in both cases are close to 0.05. These findings align with prior work in which adopting protective behaviors differed by age [38] and gender [79, 100]. Other demographic variables and email-related factors are not significantly correlated with prior awareness.

RQ3: What factors influence participants' awareness of data breaches that affected them? Participants were unaware of 74% of the breaches presented in our study, suggesting that current methods of informing consumers about data breaches might be ineffective. Prior awareness primarily came from interactions with the breached company or third-party notification services. Notably, participants were significantly more likely to be aware of older breaches. A longer time-lapse might provide participants with more opportunities to learn about the breach, and once aware, participants' memory of the breach does not seem to fade away.

5.4 RQ4: Emotional Response and Concerns towards Breaches

Participants indicated their concern using a 5-point Likert item for each shown breach (Question 15) and for each data type leaked in a breach (Question 17). We also asked participants to describe their feelings regarding the breach (Question 14, open-ended).

Quantitative ratings of concern level. Among 792 breach-specific responses, the median concern level regarding the breach was "somewhat concerned." Less than half reported either no concern (151, 19%) or being very/extremely concerned (197, 25% combined). Figure 4 shows concern levels for commonly leaked data types. Participants were most concerned about leaks of physical address (52% very/extremely), passwords (47% very/extremely), and phone number (42% very/extremely). Other leaked data types that participants felt less concerned about were employer information (38% not at all), social media profile (42% not at all), job title (46% not at all), and gender (65% not at all).

We sought to identify factors that might impact concern level through a mixed-effect linear regression on overall concern Likert responses. We included email address-related factors and prior awareness as independent variables, hypothesizing that participants would be more concerned about frequently used email addresses or if they had not been aware of a breach. We also included the number of breached data types and the breach status of data types for which more than 50% of responses were "somewhat concerned" or above in Figure 4, namely password, physical address, phone number, date of birth, IP address, and name.[7] We hypothesized that as the amount or sensitivity of leaked data types increases, the concern level would increase. Additionally, we included

[7] Email address was not included because it was exposed in all breaches in our sample, making no positive vs. negative cases.

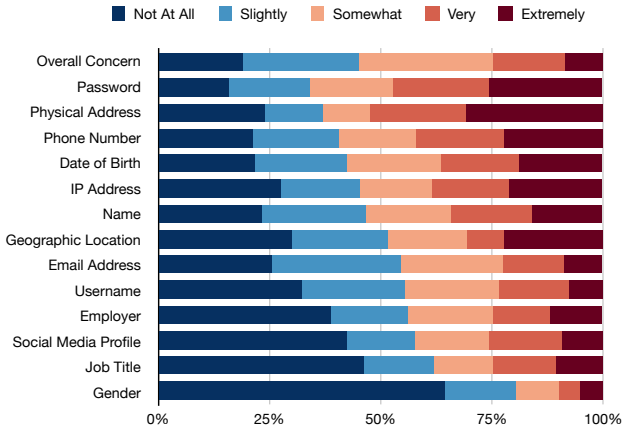

Figure 4: Overall concern (Question 15) about the breach and levels of concern for the 13 most commonly leaked information types in our sample breaches (Question 17).

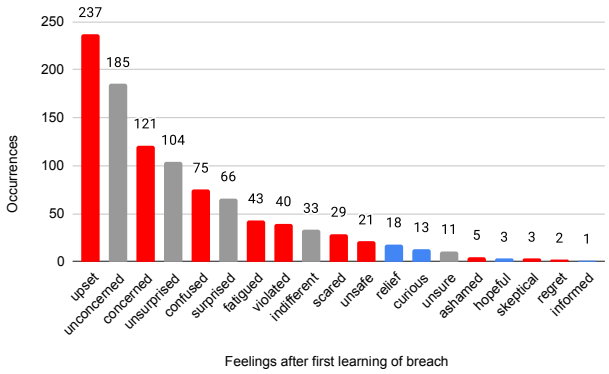

Figure 5: Code frequencies for feelings after first learning about a breach ($n = 792$); red bars indicate negative feelings, gray neutral, blue positive, according to Emolex ratings [52].

the breaches' age since participants might be more concerned about recent breaches.

The regression results do not reveal any significant factors impacting overall concern except the intercept ($b_{intercept}$=2.52, SE=.31, p<.001), indicating that participants likely default to between "slightly concerned" and "somewhat concerned." The model's $f^2 = 0.03$ indicates a small effect size. The absence of influential factors on concern may be due to data types known to trigger more concerns, such as financial information and social security numbers, being underrepresented in our sample's breaches (see Figure 2). Even relatively sensitive data types in our sample still had a fair number of "not at all/slightly concerned" responses.

Various emotions in qualitative responses. Figure 5 shows the wide range of emotions reflected in participants' open-ended responses about their feelings after learning of a breach affecting them. In 237 (30%) cases, participants reported feeling upset (including annoyed, frustrated, mad, and angry), mostly toward the breached organization. The upset came from not having been properly informed (e.g., *"I was very disappointed . . . they hid the fact that there was a data breach from everyone for three months"*), the organization's poor security measures (e.g., *"don't run an entirely online business if you cant do basic security"*), or violation of consumers' trust (e.g., *"I joined this site to read a story my granddaughter had written and thought it was completely safe"*). These emotions align with the "risk as feelings" theory, which highlights that people experience dread and outrage in comprehending risks [80], and that such affective responses greatly influence their subsequent decision-making, sometimes overriding cognitive assessments [48].

Mirroring the Likert responses, feeling unconcerned about a breach was common (185, 23%). Many participants believed that the exposed data was not sensitive (e.g., *"I had only used the free version of that site, so I had not entered any payment information"*). Others were unconcerned because they rarely interacted with nor knew the breached organization (e.g., *"I don't even know what this site is, so I don't think that them having my info . . . is a huge deal"*). Some were unconcerned due to confidence in their security habits, including regularly changing passwords (25), avoiding password reuse (10), and enabling 2FA (4). A few participants were unconcerned due to a lack of experienced impacts (e.g., *"I'm not especially worried because I haven't detected any suspicious activity"*) or optimism bias (e.g., *"I feel like a drop in the bucket since there were 711 million emails affected"*).

104 (13%) responses reported feeling unsurprised whereas 66 (8%) reported feeling surprised. Unsurprised participants explained that they never trusted the breached organization or already knew about the breach. Conversely, surprised participants stated that they had never used the breached organization's service or trusted the organization.

In another 75 (9%) cases, participants expressed confusion due to unfamiliarity with the breached organization or not remembering having an account. Other prominent emotions included fatigued (43, 5%), violated (40, 5%), indifferent (33, 4%), scared (29, 4%), unsafe (18, 2%), relieved (18, 2%), or curious about why the breach happened (13, 2%). Those who expressed fatigue stressed that breaches were inevitable (e.g., *"It's the internet and things WILL be leaked somehow, either by hackers or by incompetence at the company that is holding your information anyhow"*). This attitude is akin to the "digital resignation" phenomenon [20]: many people's inaction in the face of privacy infringements are not necessarily because they do not care, but because they are resigned and convinced that surveillance is inescapable. Notably, neutral emotions, like curiosity, or positive emotions, like relief, were rare. Participants were relieved when sensitive data like financial information was not involved or that they were now aware of the breach and could take proper action.

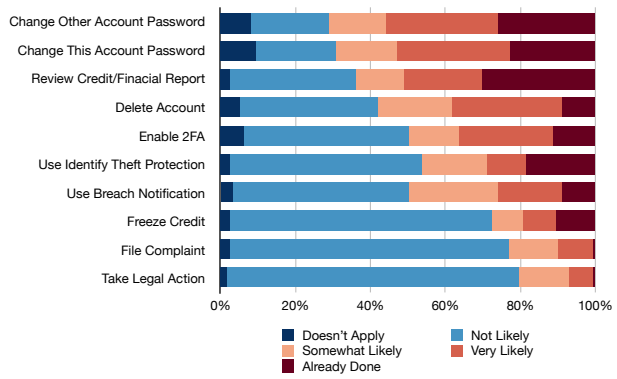

Figure 6: Intention to take actions within the next 30 days.

RQ4: What are participants' emotional responses to data breaches that affected them? While some leaked data types (e.g., password, physical address, and phone number) triggered heightened concerns, overall participants reported low concern about data breaches: 56% were slight or somewhat concerned, and 19% were not at all concerned. However, participants expressed a rich set of (mostly negative) emotions beyond concerns, such as feeling upset with the breached organization and feeling fatigued by the sheer number of data breaches nowadays.

5.5 RQ5: Behavioral Reactions to Breaches

For the 143 breaches participants were already aware of before our study, we further asked if they had taken any action in response (Questions 18). The most common action taken was to change passwords (87, 61%). 15 specified they changed the password for the breached account, and 27 mentioned changing the password across multiple accounts that might use the leaked password. Five further mentioned changing their email account's password; this could be due to a misconception that their email account, not the account with the breached organization, was compromised. Participants also described other password-related practices triggered by the breach, such as using unique passwords, using a password manager, and making passwords more complicated.

Participants reported having taken a variety of actions related to their account with the breached organization. 18 (13%) deleted or deactivated the account, and one mentioned reviewing accounts on other websites and deleting them as needed. Five mentioned enabling 2FA for the breached organizations' account, for other accounts, or for their email account. Four reported checking the breached organization's account to see if it stored any sensitive data or if there had been any suspicious activity. In 31 (22%) cases, participants reported doing nothing in reaction; the percentage was lower than that in Ponemon's 2014 survey (32%) [31], but still substantial.

Additionally, we asked all participants with at least one breach to indicate, for each breach, how likely they were to initiate ten provided actions within the next 30 days or whether they had taken action already. We only include 500 breach-specific responses in the following analysis due to a data storage issue, excluding incomplete responses. Figure 6 shows the results. Of the ten provided actions, changing the password for the breached organizations' account or other accounts were the most popular, receiving more than half of likely/already done responses. "Review credit reports and/or financial statements" had the highest percentage of already done (30%). By contrast, most participants selected "not likely" for four actions — "use a credit/identity monitoring service," "place a credit freeze on my credit reports," "file a complaint with a consumer protection agency," and "take legal action against the breached organization." This finding is understandable given that most leaked data types such as email addresses and passwords are considered "non-sensitive records" according to ITRC's report [30].

We sought to understand factors that would impact the likelihood of having taken any of the ten provided actions through a mixed-effect logistic regression. For independent variables, we discarded variables related to email habits since many of the listed actions were unrelated to one's email account. We kept all other independent variables from the concern regression model, namely prior awareness, the breach's age, the number of breached data types, and the breach status of six data types with relatively high concern levels. We further included overall concern Likert responses as an independent variable. Results in Table 5 show a significant intercept, indicating that participants were likely to default

Table 5: Logistic regression on taking actions.

	Est.	OR	95% CI	p-value
(Intercept)	-3.27	0.04	$[0.002, 0.61]$	.02
Awareness yes (vs. no)	5.97	390.48	$[45.72, 3334.79]$	< 0.001
Breach age years	-0.03	0.97	$[0.77, 1.21]$	.77
Num. of types numeric	.12	1.13	$[0.85, 1.50]$	.39
Password yes (vs. no)	-0.18	0.84	$[0.18, 3.79]$	.82
Physical Addr. yes (vs. no)	-0.26	0.77	$[0.16, 3.71]$	.75
Phone Num. yes (vs. no)	-0.29	0.75	$[0.19, 3.02]$	.69
Date of birth yes (vs. no)	-0.24	0.79	$[0.17, 3.62]$	.76
IP Addr. yes (vs. no)	-0.20	0.82	$[0.26, 2.64]$	.74
Name yes (vs. no)	-0.19	0.83	$[0.21, 3.22]$	.79
Concern numeric	0.80	2.22	$[1.28, 3.86]$	.005

to inaction with no leaked data and no prior awareness or concern ($OR_{intercept}$=0.04, p=.02). Being aware of a breach significantly increased the likelihood of having taken any of the listed actions (OR^{no}_{yes}=390.48, p<.001). This is unsurprising given that participants who were unaware of being affected had little motivation to engage in protective measures. Additionally, more concern was significantly correlated with a higher likelihood of having taken action: for a one-unit increase of concern on the 5-point Likert scale, the odds of having taken action increase by 2.22 ($OR_{concern}$=2.22, p=.005).

RQ5: What factors influence participants' likelihood to take action in response to data breaches that affected them? Participants' intention to act varies among protective measures: they were more amenable to change passwords and check credit reports/financial records than other actions. The regression results reveal that awareness and concern drive the likelihood of taking action, while other factors such as the leaked data types do not impact the outcome. Our findings suggest that to motivate consumers to react to breaches, they must first be aware that the breach occurred and feel concerned enough to invest in mitigation efforts.

6 Discussion

We examined individuals' awareness, perception, and responses to specific data breaches that had exposed their email addresses and other information. Compared to RAND's 2016 survey [1], in which 44% reported already knowing about a breach before receiving a notification, participants' prior awareness was much lower in our sample. This finding is concerning as our results suggest that unawareness creates a substantial barrier for taking mitigating action. Participants also reported a lower level of overall concern than in prior work [31, 37]: this might result from a methodological difference, as our participants reflected on specific breaches affecting them rather than on breaches in general [1, 31] or on hypothetical scenarios [37]. Another possible reason is that the leaked data types in the HIBP database are mostly categorized as non-sensitive records [30]. While participants named potential consequences of data breaches such as more spams and increased risks of identity theft, similar to prior work [37, 99], many considered these events would have little to no impact on their lives. Most participants also exhibited misconceptions about what led to themselves being affected by breaches, blaming their own email or password behaviors rather than the breached organization.

Set stricter legal requirements for notifying consumers. Our study reflects a sad reality that many individuals are unaware that they are affected by breaches, at least for breaches exposing email addresses. Current breach notification requirements, mechanisms, and tools fail to reach data breach victims. Nonetheless, awareness was a crucial trigger of taking action, according to our regression results.

Stricter regulatory requirements may help establish high standards for breach notifications, which in turn raise awareness. Simply requiring companies to send the notification is not enough as the notification also needs to be effective [8, 98]. For instance, prior work highlights the role of media reports in informing and shaping attitudes of data breaches [1, 15]. Our findings indicate that notifications from breached organizations or third-party services are more relevant. Given that individuals may not stick with one channel to learn about breaches, breached organizations could be mandated to notify consumers in multiple channels instead of the most convenient one, and obtain confirmation from victims that the notification was received. Regarding when to notify, Art. 34 of Europe's General Data Protection Regulation (GDPR) specifies that consumer-facing notifications are only needed for breaches that "result in a high risk" to data subjects [22]. We argue that this should be done for *all* breaches, given that many court cases struggle to assess risks and harms caused by data breaches [81]; this requirement would also be more in line with consumer preferences [54]. Alternatively, less ambiguous criteria should be set for high-risk breaches, e.g., in California, consumer-facing notifications are mandated when the breach involves unencrypted personally identifiable information [82].

Use novel approaches in notifying consumers. Prior research on SSL warnings [3, 23, 24] shows that in-browser warnings effectively raise threat awareness and encourage safer practices. Similarly, data breach notifications could explore approaches beyond letters and emails, such as in-situ methods whereby visiting affected sites leads to a notification [17], as recently pursued by some browsers and password managers that warn users if saved passwords appeared in credential dumps [44, 62].

Notifications should also consider non-adherence: among participants who were already aware of a breach before our study, 22% reported doing nothing in response to that breach; emotions like fatigue and resignation were also noted. Drawing from warning design literature on mitigating fatigue in email-based notifications [7, 42], one could build systems that highlight unread breach notifications in email clients, similar to Gmail's reminders to reply to emails [10]. The contents of such emails could also be automatically parsed and reformatted to guide attention to important details.

Address misconceptions. Participants commonly blamed their own email habits or security practices for data breaches, and such misconceptions exacerbate a power asymmetry — rather than demanding that organizations improve security measures or that regulators hold them accountable, participants blamed themselves. Consumers should be reminded that the root cause of breaches is security issues in the breached

organization, and there are actions that can hold the breached organization accountable, such as filing a complaint with a consumer protection agency (e.g., the Federal Trade Commission for US breaches).

Participants also differed regarding perceived impacts of breaches. Those who had not experienced adverse impacts mostly did not take data breaches seriously. Conversely, those who had experienced an adverse event reported emotional distress and resulting behavioral changes. Indeed, not everyone would experience the negative consequences of not reacting to data breaches, but the cost is real and immediate when the consequences manifest. Breach notifications and education materials should stress that good security practices, such as using unique passwords and 2FA, can dampen the severity of a breach's impact even though they do not decrease one's likelihood of being affected by a breach. While these precautionary measures might not provide instant gratification, they could be worthy investments considering the substantial hassles and trauma in recovering from identity theft [43] or other repercussions of breaches.

Develop tools to help consumers react to breaches. While consumers may not be able to prevent breaches from occurring, actions are available for mitigating the aftermath of a breach. Our findings show that some straightforward actions, such as changing passwords, had high adoption rates or intention to adopt. Yet, the majority of provided actions were much less popular (see Figure 6), indicating the need to offer more relevant and usable protective measures to affected individuals.

One of our key findings is that extensive use of an email account (e.g., use it for a long time and check it frequently) significantly increased the email address's likelihood of being involved in a breach. Yet, simply asking users to reduce their usage or abandon their email account is not a viable solution, as it also diminishes the email account's utility. Instead, drawing from some participants' descriptions of creating dedicated email accounts for registration on low-value sites, we see the promise of more automated tools to offer unique email aliases for account registration. Such features could further be integrated into other technologies with broader adoption, such as browsers or password managers, to create a more streamlined experience (e.g., through auto-filling). Recent respective efforts include "Sign in with Apple"[8] and "Firefox Relay"[9], both of which support the generation of a unique, random email address during account registration, which is forwarded to a user's real inbox. However, both products are currently limited to their respective ecosystems. The effectiveness, awareness, and adoption of such tools, as well as how individuals manage multiple email aliases in general, are open questions for future research.

[8] https://support.apple.com/en-us/HT210318

[9] https://blog.mozilla.org/firefox/firefox-relay/

Increasing responsibilities of breached organizations. Our participants exhibited a low awareness of data breaches, which in turn serves as a precursor to the low intention for certain protective measures. This lack of awareness and self-protection among participants indicates that breached organizations should play a more active role in protecting affected individuals. Notifying victims should not absolve breached organizations from further responsibility — they should further ensure that consumers have viable remediation solutions and assist in the recovery process, such as offering support in identity restoration. Rather than defaulting to conventional credit and identity monitoring services, which are known to provide little preventative protection [40], breached organizations could offer victims email alias generators, password managers, or other more promising mitigation tools by partnering with respective service providers. Regulators should also set and frequently revisit requirements for the types of services breached organizations must offer as compensation.

Importantly, breached organizations have financial incentives for transparent post-breach communications and active mitigation. Prior work shows that data breach notifications provide a venue for impression management and repairing damaged trust [33]. Moreover, breached organizations that provide affected individuals with free credit monitoring services face a lower likelihood of lawsuits [73]. Regulators should also create meaningful incentives for organizations to act accordingly. For instance, the GDPR's threat of substantial fines has resulted in a heightened effort by organizations worldwide to overhaul their privacy and security programs.

7 Conclusion

Our study provides insights into individuals' awareness, perception, and responses to data breaches. We applied a novel method that presented participants with specific data breaches exposing their email addresses and other information. Our findings reveal some concerning aspects, such as participants' low awareness of breaches that affected them and misconceptions about the causes and impacts of being involved in these breaches. We outline potential avenues for addressing these identified issues — improving consumers' awareness of breaches affecting them, developing novel and useful tools to help consumers mitigate the impacts of breaches, and increasing the responsibility and accountability of breached organizations.

Acknowledgements

This research was partially supported by a NortonLifeLock Graduate Fellowship and the Helmholtz Association (HGF) through the subtopic Engineering Secure Systems (ESS).

References

[1] Lillian Ablon, Paul Heaton, Diana Catherine Lavery, and Sasha Romanosky. Consumer attitudes toward data breach notifications and loss of personal information. Technical report, Rand Corp., 2016.

[2] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Wang Yang, and Shomir Wilson. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys*, 50(3):1–41, 2017.

[3] Devdatta Akhawe and Adrienne Porter Felt. Alice in warningland: A large-scale field study of browser security warning effectiveness. In *USENIX Security Symp.*, pages 257–272, 2013.

[4] Mohamed Alsharnouby, Furkan Alaca, and Sonia Chiasson. Why phishing still works: User strategies for combating phishing attacks. *Intl. J. of Human-Computer Studies*, 82:69–82, 2015.

[5] J. Craig Anderson. Identity theft growing, costly to victims, 2013. https://www.usatoday.com/story/money/personalfinance/2013/04/14/identity-theft-growing/2082179/.

[6] Julio Angulo and Martin Ortlieb. "wth..!?!" experiences, reactions, and expectations related to online privacy panic situations. In *Symp. On Usable Privacy and Security*, pages 19–38, 2015.

[7] Eric Bachura, Rohit Valecha, Rui Chen, and Raghav H Rao. Modeling public response to data breaches. In *Americas Conf. on Info. Sys.* AIS eLibrary, 2017. Art. 43.

[8] Fabio Bisogni. Proving limits of state data breach notification laws: Is a federal law the most adequate solution? *J. of Info. Policy*, 6(1):154–205, 2016.

[9] Cristian Bravo-Lillo, Lorrie Faith Cranor, Julie Downs, and Saranga Komanduri. Bridging the gap in computer security warnings: A mental model approach. *IEEE Security & Privacy*, 9(2):18–26, 2010.

[10] Scott Brown. Did you forget to reply to an email? the new gmail will remind you, 2018. https://www.androidauthority.com/gmail-nudges-feature-865435/.

[11] Mohamad Adam Bujang, Nadiah Sa'at, Tg Mohd Ikhwan Tg Abu Bakar, et al. Sample size guidelines for logistic regression from observational studies with large population: emphasis on the accuracy between statistics and parameters based on real life clinical data. *The Malaysian J. of Medical Sciences*, 25(4):122, 2018.

[12] Hsiangting Shatina Chen and Tun-Min Jai. Trust fall: data breach perceptions from loyalty and non-loyalty customers. *The Service Industries J.*, pages 1–17, 2019.

[13] CNBC. Target gives 10% discount to shoppers after data breach, 2013. https://www.cnbc.com/2013/12/20/target-gives-10-discount-to-shoppers-after-data-breach.html.

[14] Sauvik Das, Laura A Dabbish, and Jason I Hong. A typology of perceived triggers for end-user security and privacy behaviors. In *Symp. on Usable Privacy and Security*, pages 97–115, 2019.

[15] Sauvik Das, Joanne Lo, Laura Dabbish, and Jason I Hong. Breaking! A typology of security and privacy news and how it's shared. In *ACM CHI Conf. on Human Factors in Computing Sys.*, 2018. Art. 1.

[16] Behnam Dayanim and Edward George. Data breach litigation and regulatory enforcement: A survey of our present and how to prepare for the future. *Cyber Security*, 1(4):301–315, 2018.

[17] Joe DeBlasio, Stefan Savage, Geoffrey M Voelker, and Alex C Snoeren. Tripwire: Inferring internet site compromise. In *ACM SIGCOMM Conf. on Internet Measurement*, pages 341–354, 2017.

[18] Digital Shadows Photon Research Team. From exposure to takeover: The 15 billion stolen credentials allowing account takeover. Technical report, 2019.

[19] Alexandra Dmitrienko, Christopher Liebchen, Christian Rossow, and Ahmad-Reza Sadeghi. On the (in) security of mobile two-factor authentication. In *Intl. Conf. on Financial Cryptography and Data Security*, pages 365–383, 2014.

[20] Nora A Draper and Joseph Turow. The corporate cultivation of digital resignation. *New Media & Society*, 21(8):1824–1839, 2019.

[21] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *ACM CHI Conf. on Human Factors in Computing Sys.*, pages 1065–1074, 2008.

[22] European Parliament. Regulation (eu) 2016/679 of the european parliament and of the council, 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

[23] Adrienne Porter Felt, Alex Ainslie, Robert W Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. Improving SSL warnings: Comprehension and adherence. In *ACM CHI Conf. on Human Factors in Computing Sys.*, pages 2893–2902, 2015.

[24] Adrienne Porter Felt, Robert W Reeder, Hazim Almuhimedi, and Sunny Consolvo. Experimenting at scale with google chrome's SSL warning. In *ACM CHI Conf. on Human Factors in Computing Sys.*, pages 2667–2670, 2014.

[25] Gemalto. Data breaches & customer loyalty 2017. Technical report, 2018.

[26] Google. Cleaning up after password dumps, 2014. https://security.googleblog.com/2014/09/cleaning-up-after-password-dumps.html.

[27] Claire Greene and Joanna Stavins. Did the target data breach change consumer assessments of payment card security? *J. of Payments Strategy & Sys.*, 11(2):121–133, 2017.

[28] Zahra Hassanzadeh, Sky Marsen, and Robert Biddle. We're here to help: Company image repair and user perception of data breaches. In *ACM Graphics Interface Conf.*, 2020.

[29] Troy Hunt. Have i been pwned: Check if you have an account that has been compromised in a data breach, 2020. https://haveibeenpwned.com/.

[30] Identity Theft Resource Center. Data breach report. Technical report, 2020.

[31] Ponemon Institute. The aftermath of a data breach: Consumer sentiment. Technical report, 2014.

[32] Markus Jakobsson. Two-factor inauthentication–the rise in sms phishing attacks. *Computer Fraud & Security*, 2018(6):6–8, 2018.

[33] Alexander Jenkins, Murugan Anandarajan, and Rob D'Ovidio. 'all that glitters is not gold': The role of impression management in data breach notification. *Western J. of Comm.*, 78(3):337–357, 2014.

[34] Jeffrey Jenkins, Alexandra Durcikova, and Jay F Nunamaker Jr. Mitigating the Security Intention-Behavior Gap: The Moderating Role of Required Effort on the Intention-Behavior Relationship. *J. of the Assoc. for Info. Sys.*, 22(1):1, 2021.

[35] Rhoda C Joseph. Data breaches: Public sector perspectives. *IT Professional*, 20(4):57–64, 2017.

[36] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. "My Data Just Goes Everywhere:" user mental models of the internet and implications for privacy and security. In *Symp. On Usable Privacy and Security*, pages 39–52, 2015.

[37] Sowmya Karunakaran, Kurt Thomas, Elie Bursztein, and Oxana Comanescu. Data breaches: User comprehension, expectations, and concerns with handling exposed data. In *Symp. on Usable Privacy and Security*, pages 217–234, 2018.

[38] Murat Kezer, Barış Sevi, Zeynep Cemalcilar, and Lemi Baruh. Age differences in privacy attitudes, literacy and privacy management on facebook. *Cyberpsychology*, 10(1), 2016.

[39] Bokyung Kim, Kristine Johnson, and Sun-Young Park. Lessons from the five data breaches: Analyzing framed crisis response strategies and crisis severity. *Cogent Business & Mgmt.*, 4(1):1–17, 2017.

[40] Brian Krebs. Are credit monitoring services worth it?, 2014. https://krebsonsecurity.com/2014/03/are-credit-monitoring-services-worth-it/.

[41] Thomas Kude, Hartmut Hoehle, and Tracy Ann Sykes. Big data breaches and customer compensation strategies: Personality traits and social influence as antecedents of perceived compensation. *Intl. J. of Operations & Production Mgmt.*, 37(1):56–74, 2017.

[42] Juhee Kwon and M Eric Johnson. The market effect of healthcare security: Do patients care about data breaches? In *Workshop on the Economics of Info. Security*, 2015.

[43] Charity Lacey. The aftermath: the non-economic impacts of identity theft. Technical report, Identity Theft Resource Center, 2018.

[44] Ravie Lakshmanan. Chrome and firefox will now alert you about data breaches involving your accounts, 2019. https://thenextweb.com/security/2019/10/23/chrome-and-firefox-will-now-alert-you-about-data-breaches-involving-your-accounts/.

[45] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174, 1977.

[46] Elmer Lastdrager, Inés Carvajal Gallardo, Pieter Hartel, and Marianne Junger. How effective is anti-phishing training for children? In *Symp. on Usable Privacy and Security*, pages 229–239, 2017.

[47] Ron Lieber. How to protect yourself after the equifax breach, 2019. https://www.nytimes.com/interactive/2017/your-money/equifax-data-breach-credit.html.

[48] George F Loewenstein, Elke U Weber, Christopher K Hsee, and Ned Welch. Risk as feelings. *Psychological bulletin*, 127(2):267, 2001.

[49] Rebecca T Mercuri. Analyzing security costs. *Comms. of the ACM*, 46(6):15–18, 2003.

[50] Vyacheslav Mikhed and Michael Vogan. Out of sight, out of mind: consumer reaction to news on data breaches and identity theft. *FRB of Philadelphia Working Paper*, pages 15–42, 2015.

[51] Vyacheslav Mikhed and Michael Vogan. How data breaches affect consumer credit. *J. of Banking & Finance*, 88:192–207, 2018.

[52] Saif M. Mohammad and Peter D. Turney. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465, 2013.

[53] Mozilla. Firefox monitor, 2020. https://monitor.firefox.com/.

[54] Patrick Murmann, Delphine Reinhardt, and Simone Fischer-Hübner. To be, or not to be notified. In *Int. Conf. on ICT Systems Security and Privacy Protection*, pages 209–222. Springer, 2019.

[55] Steven Muzatko and Gaurav Bansal. Timing of data breach announcement and e-commerce trust. In *Midwest Assoc. for Info. Sys. Conf.*, 2018. Art. 7.

[56] Donald A Norman. Some observations on mental models. In Dedre Gentner and Albert L Stevens, editors, *Mental models*, chapter 1, pages 7–14. Hillsdale, 1983.

[57] Stefan Palan and Christian Schitter. Prolific.ac – a subject pool for online experiments. *J. of Behavioral and Experimental Finance*, 17:22–27, 2018.

[58] Sarah Pearman, Jeremy Thomas, Pardis Emami Naeini, Hana Habib, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, and Alain Forget. Let's go in for a closer look: Observing passwords in their natural habitat. In *ACM Conf. on Computer and Comms. Security*, pages 295–310, 2017.

[59] Peng Peng, Chao Xu, Luke Quinn, Hang Hu, Bimal Viswanath, and Gang Wang. What happens after you leak your password: Understanding credential sharing on phishing sites. In *ACM Asia Conf. on Computer and Comms. Security*, pages 181–192, 2019.

[60] Justin Petelka, Yixin Zou, and Florian Schaub. Put your warning where your link is: Improving and evaluating email phishing warnings. In *ACM CHI Conf. on Human Factors in Computing Sys.*, 2019. Art. 518.

[61] Rachael M Peters. So you've been notified, now what: The problem with current data-breach notification laws. *Arizona Law Rev.*, 56:1171–1202, 2014.

[62] Katie Petrillo. Protect your accounts with breach alerts through lastpass, 2018. https://blog.lastpass.com/2018/11/protect-your-accounts-with-breach-alerts-through-lastpass/.

[63] Privacy Rights Clearinghouse. Data breaches, 2020. https://privacyrights.org/data-breaches.

[64] Emilee Rader, Samantha Hautea, and Anjali Munasinghe. "I Have a Narrow Thought Process": Constraints on explanations connecting inferences and self-perceptions. In *Symp. on Usable Privacy and Security*, pages 457–488, 2020.

[65] Emilee Rader and Rick Wash. Identifying patterns in informal sources of security information. *J. of Cybersecurity*, 1(1):121–144, 2015.

[66] Lee Rainie, Sara Kiesler, Ruogu Kang, Mary Madden, Maeve Duggan, Stephanie Brown, and Laura Dabbish. Anonymity, privacy, and security online. Technical report, Pew Research Center, 2013.

[67] Elissa M Redmiles, Amelia R Malone, and Michelle L Mazurek. I think they're trying to tell me something: Advice sources and selection for digital security. In *IEEE Symp. on Security and Privacy*, pages 272–288, 2016.

[68] Elissa M Redmiles, Noel Warford, Amritha Jayanti, Aravind Koneru, Sean Kross, Miraida Morales, Rock Stevens, and Michelle L Mazurek. A comprehensive quality evaluation of security and privacy advice on the web. In *USENIX Security Symp.*, pages 89–108, 2020.

[69] Robert W Reeder, Iulia Ion, and Sunny Consolvo. 152 simple steps to stay safe online: security advice for non-tech-savvy users. *IEEE Security & Privacy*, 15(5):55–64, 2017.

[70] William Roberds and Stacey L Schreft. Data breaches and identity theft. *J. of Monetary Economics*, 56(7):918–929, 2009.

[71] Steve Roberts. Learning lessons from data breaches. *Network Security*, 2018(11):8–11, 2018.

[72] Sasha Romanosky. Examining the costs and causes of cyber incidents. *J. of Cybersecurity*, 2(2):121–135, 2016.

[73] Sasha Romanosky, David Hoffman, and Alessandro Acquisti. Empirical analysis of data breach litigation. *J. of Empirical Legal Studies*, 11(1):74–104, 2014.

[74] Sasha Romanosky, Rahul Telang, and Alessandro Acquisti. Do data breach disclosure laws reduce identity theft? *J. of Policy Analysis and Mgmt.*, 30(2):256–286, 2011.

[75] Johnny Saldaña. *The Coding Manual for Qualitative Researchers*. Sage, 2015.

[76] Robert Schoshinski. Equifax data breach: Pick free credit monitoring, 2019. https://www.consumer.ftc.gov/blog/2019/07/equifax-data-breach-pick-free-credit-monitoring.

[77] Richard Shay, Iulia Ion, Robert W Reeder, and Sunny Consolvo. "My religious aunt asked why I was trying to sell her viagra": experiences with account hijacking. In *ACM CHI Conf. on Human Factors in Computing Sys.*, pages 2657–2666, 2014.

[78] Paschal Sheeran and Thomas L Webb. The intention–behavior gap. *Social and Personality Psych. Compass*, 10(9):503–518, 2016.

[79] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions. In *ACM CHI Conf. on Human Factors in Computing Sys.*, pages 373–382, 2010.

[80] Paul Slovic, Melissa L Finucane, Ellen Peters, and Donald G MacGregor. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, 24(2):311–322, 2004.

[81] Daniel J Solove and Danielle Keats Citron. Risk and anxiety: A theory of data-breach harms. *Texas Law Rev.*, 96:737–786, 2017.

[82] State of California. California Civil Code 1798.82. https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CIV&sectionNum=1798.82.

[83] Rahul Telang. Policy framework for data breaches. *IEEE Security & Privacy*, 13(1):77–79, 2015.

[84] The Federal Trade Commission. Credit Freeze FAQs, 2019. https://www.consumer.ftc.gov/articles/0497-credit-freeze-faqs.

[85] The Federal Trade Commission. When information is lost or exposed, 2020. https://www.identitytheft.gov/databreach.

[86] The Firefox Frontier. What to do after a data breach, 2019. https://blog.mozilla.org/firefox/what-to-do-after-a-data-breach/.

[87] Kurt Thomas, Frank Li, Ali Zand, Jacob Barrett, Juri Ranieri, Luca Invernizzi, Yarik Markov, Oxana Comanescu, Vijay Eranti, Angelika Moscicki, et al. Data breaches, phishing, or malware? understanding the risks of stolen credentials. In *ACM Conf. on Computer and Comms. Security*, pages 1421–1434, 2017.

[88] Dana Turjeman and Fred M Feinberg. When the data are out: Measuring behavioral changes following a data breach. *SSRN Electronic J.*, 2019.

[89] Jennifer R Veltsos. An analysis of data breach notifications as negative news. *Business Comm. Quarterly*, 75(2):192–207, 2012.

[90] Paul Wagenseil. What to do after a data breach, 2019. https://www.tomsguide.com/us/data-breach-to-dos,news-18007.html.

[91] Rick Wash. Folk models of home computer security. In *Symp. on Usable Privacy and Security*, 2010. Art. 11.

[92] Rick Wash, Emilee Rader, Ruthie Berman, and Zac Wellmer. Understanding password choices: How frequently entered passwords are re-used across websites. In *Symp. on Usable Privacy and Security*, pages 175–188, 2016.

[93] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L Mazurek, and Blase Ur. Oh, the places you've been! User reactions to longitudinal transparency about third-party web tracking and inferencing. In *ACM Conf. on Computer and Comms. Security*, pages 149–166, 2019.

[94] Kimberly A Whitler and Paul W Farris. The impact of cyber attacks on brand image: Why proactive marketing expertise is needed for managing data breaches. *J. of Advertising Research*, 57(1):3–9, 2017.

[95] Victoria Woollaston. Facebook and netflix reset passwords after data breaches, 2016. https://www.wired.co.uk/article/facebook-netflix-password-reset.

[96] Achim Zeileis, Christian Kleiber, and Simon Jackman. Regression models for count data in R. *J. of Stat. Software*, 27(8):1–25, 2008.

[97] Verena Zimmermann, Merve Bennighof, Miriam Edel, Oliver Hofmann, Judith Jung, and Melina von Wick. 'Home, Smart Home' – exploring end users' mental models of smart homes. *Usable Security and Privacy Workshop at Mensch und Computer*, 2018.

[98] Yixin Zou, Shawn Danino, Kaiwen Sun, and Florian Schaub. You 'might' be affected: An empirical analysis of readability and usability issues in data breach notifications. In *ACM CHI Conf. on Human Factors in Computing Sys.*, 2019. Art. 194.

[99] Yixin Zou, Abraham H Mhaidli, Austin McCall, and Florian Schaub. "I've got nothing to lose": Consumers' risk perceptions and protective actions after the equifax data breach. In *Symp. on Usable Privacy and Security*, pages 197–216, 2018.

[100] Yixin Zou, Kevin Roundy, Acar Tamersoy, Saurabh Shintre, Johann Roturier, and Florian Schaub. Examining the adoption and abandonment of security, privacy, and identity theft protection practices. In *ACM CHI Conf. on Human Factors in Computing Sys.*, 2020. Art. 443.

Appendix

A Survey Material

A.1 Informed consent

Study Title: Awareness, Risk Perception, and Reaction Toward Data Breaches

Principal Investigators: *REDACTED*

Purpose of this Study: We are conducting a research study to understand how users perceive and react to data breaches.

Description of your involvement: If you agree to be part of the research study, we will ask you to complete an online survey where you will be asked to review data breach records associated with one of your email addresses based on a public database of security breaches (haveibeenpwned.com) and answer a few questions about the displayed records. We anticipate the survey will take about 15 minutes.

Requirements: To participate in the study, you must (1) be 18 years old or older; and (2) currently live in the United States.

Benefits: You may not receive a direct benefit from participating, but this study will help us develop better systems and technologies that empower Internet users to protect themselves against data breaches.

Risks: The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or during use of the Internet.

Compensation: You will be compensated $2.50 upon completing the survey.

Confidentiality: By participating in the study, you understand and agree that the *REDACTED* may be required to disclose your consent form, data and other personally identifiable information as required by law, regulation, subpoena or court order. Otherwise, your confidentiality will be maintained in the following manner:

Your data and consent form will be kept separate. Your research data will be stored securely and will only be accessible to the study team. By participating, you understand and agree that the data and information gathered during this study may be published in an academic journal or conference paper. You will not be asked to provide any direct personal identifiers in the study apart from your email address. *We do not track or store your email address as part of this study, and we will not be able tie your email address to any results or analysis. All records of your email address will reside only in temporary storage to facilitate the lookup of data breaches your email address was involved in and will be deleted following the completion of this task. The researchers will never see your email address.*

Right to Ask Questions & Contact Information: If you have questions about this research, you may contact the study team at *REDACTED*

The *REDACTED* Institutional Review Board has determined that this study is exempt from IRB oversight.

Voluntary Consent: By proceeding to the next page, you are agreeing to participate in this study. Please be sure that we have answered any questions you may have about the study, and you understand what you are being asked to do. You may contact the researchers at any time by emailing *REDACTED* if you think of a question later.

STATEMENT BY PERSON AGREEING TO PARTICIPATE IN THIS STUDY I have read this informed consent document and the material contained in it has been explained to me. I understand each part of the document, all my questions have been answered, and I freely and voluntarily choose to participate in this study. I can choose to withdraw from this research project at any time without penalty.

A.2 Email address-related questions

We are going to ask you to enter your most commonly used email address at the bottom of this page. We will use your email address to look up whether your email address has been disclosed in any data breaches (also called "security breaches"), using the public lookup service for data breaches haveibeenpwned.com. If your email address was involved in any data breaches, we will ask you some questions about those breaches.

Privacy Notice: We do not track or store your email address as part of this study, and we will not be able tie your email address to any results or analysis. All records of your email address will reside only in temporary storage to facilitate the lookup of data breaches your email address was involved in and will be deleted following the completion of this task. The researchers will never see your email address.

To access information about breaches, your email address will be communicated to haveibeenpwned.com, a public service not operated by us, which maintains a database of data breaches involving email addresses. Communication with haveibeenpwned.com will occur on secure and encrypted channels, and haveibeenpwned.com also does not permanently store email addresses used in queries. As described in their privacy policy: "Searching for an email address only ever retrieves the address from storage then returns it in the response, the searched address is never explicitly stored anywhere."

If you have any further concerns about providing your email address, you may opt-out of the survey at this time. We will remove any record of your participation. Note that if you choose to opt out, you will not be compensated.

1. Please enter your most commonly used email address. After the task, you may search for another email address, but for now, we are primarily interested in breaches that may have involved your most commonly used email address. [free text]
2. Thank you for providing your email address. Please tell us more about this email address. Whose email address is it? ○ It is my own account / I have sole ownership of this account ○ It is my shared account / I share the account with someone else (e.g., a partner or family member) ○ It is someone else's account / someone else has sole ownership of this account ○ I made up an email address just for this study
3. How often do you check emails in this account? ○ Every day ○ A few times a week ○ A few times a month ○ A few times a year
4. What do you use this email account for? Choose all that apply. ○ For professional correspondence (e.g., with colleagues, business partners) ○ For personal correspondence (e.g., friends and family members) ○ Account creation / signup for sensitive accounts (e.g., banking, taxes, etc.) ○ Account creation / signup of medium sensitive accounts (e.g., social media, online shopping) ○ Account creation / signup for low value accounts (I used it when I'm prompted to sign up but don't really care) ○ Other [free-text]
5. Approximately for how long have you been using this email account? [number entry] ○ year(s) ○ month(s) ○ week(s) ○ day(s)
6. How many other email addresses/accounts do you regularly use? (Not counting the one you entered) [number entry]

A.3 Breach-related questions

(if email not involved in a data breach) **Your email address has not been part of any of the data breaches recorded by haveibeenpwned.com.** That is great news for you, but we still would like to ask you some further questions.

7. In your opinion, what might be reasons that your email address has not been part of any data breach? [free text]
8. Do you believe another email address that you regularly use is more likely to have breaches? [yes/no]
9. Would you like to take this survey with that email address instead? [yes/no] *(if yes return participant to questions in Appendix A.2, if no continue to demographic questions in Appendix A.4)*

(if email involved in a data breach) **Your email address was part of a data breach:** According to haveibeenpwned.com your email address was part of one or more data breaches.

10. In your opinion, what might be reasons that your email address has been part of data breaches? [free text]

We will now ask you questions about three of these breaches. We will show you the full data breach history for your email address at the end of the survey.

(for up to three data breach, the following . . .)
Your email address was part of the following breach

[img and description of breach (see Figure 1)]

Please make sure you read the description of this breach, since we will now ask you a few questions with respect to this breach (the description of the breach will be available to you while answering the questions).

11. In your opinion, what might be reasons that your email address has been part of data breaches? [free text]
12. Prior to this study, were you aware that you are affected by this breach? ○ yes ○ no ○ unsure
13. *(if yes aware)* How did you first become aware that you are affected by this breach? ○ I was notified by the breached company. ○ I was notified by my bank or credit card company. ○ I was notified by a third-party breach notification service (e.g., Have I Been Pwned, Firefox Monitor, Breach Clarity). ○ I was notified by my credit monitoring or identity theft monitoring service (e.g., LifeLock, Credit Karma). ○ Someone else (e.g., a romantic partner or a family member) told me about it. ○ I found out myself through negative events in real life (e.g., suspicious activity on my credit card, locked out of online accounts.) ○ I learned about the breach through news media. ○ I do not remember. ○ Other [free text]
14. *(if yes aware)* Please describe how you felt when you learned that your information was part of this breach
    (if no/unsure aware) Please describe how you feel after now learning that your information was part of this breach. [free text]
15. *(if yes aware)* How concerned were you when you learned that your information was part of this breach?
    (if no/unsure aware) How concerned are you after now learning that your information was part of this breach? ○ Not at all concerned ○ Slightly concerned ○ Somewhat concerned ○ Very concerned ○ Extremely concerned
16. *(if yes aware)* Please describe how you think this breach has or will impact your life. If you suspect or have experienced impacts resulting from this breach, please describe them.
    (if no/unsure aware) Please describe how you think this breach will impact your life. If you suspect or have experienced impacts resulting from this breach, please describe them as well. [free text]
17. How concerned are you about the following data being compromised in this breach? [for each data type in the breach as provided by HIBP] ○ Not at all concerned ○ Slightly concerned ○ Somewhat concerned ○ Very concerned ○ Extremely concerned ○ I don't know ○ Does not apply to me (the company does not have my real information)
18. What did you do, if anything, after learning that your information was part of this breach? Please explain why. [free text]
19. Regarding this specific breach, please select how likely you are to initiate each the of the following actions within the next 30 days, or whether you have taken the action already. ○ Not likely ○ Somewhat likely ○ Very likely ○ I did/do this already ○ This does not apply to me / I don't understand
    (For each of the following actions:) • Change the password of my account for the breached company, if it exists • Change the password of other accounts that used the same password • Delete or deactivate my account for the breached company, if it exists • Enable two-factor authentication on my account for the breached company, if it is available • Use a credit or identity monitoring service (e.g., LifeLock, Identity Guard, IdentityForce, Credit Karma, Credit Sesame) • Use a breach notification service (e.g., Firefox Monitor, Breach Clarity, Have I Been Pwned) • Take legal action against the breached company • Review my credit reports and/or, bank/credit card statements for suspicious activity • File a complaint against the breached company with a consumer protection agency (e.g., FTC, CFPB, State Attorney General) • Place a credit freeze on my credit reports

20. Are there any other actions you would like to initiate within the next 30 days or other actions you have already taken? [free text]

A.4 Demographics & attention checks

21. Which of the following breaches were you asked about in this study? [multiple choice of the correct answer and four decoys]
22. What is your age? ○ 18-24 ○ 25-29 ○ 30-34 ○ 35-39 ○ 40-44 ○ 45-49 ○ 50-54 ○ 54-59 ○ 60-64 ○ 65+ ○ Prefer not to say
23. What is your gender? ○ Man ○ Woman ○ Non-Binary ○ Prefer not to answer ○ Other [free text]
24. What is the highest level of education you have completed? ○ Less than high school ○ High school or equivalent ○ Some college, no degree ○ Associate's degree, occupational ○ Associate's degree, academic ○ Bachelor's Degree ○ Master's Degree ○ Professional degree ○ Doctoral degree ○ Prefer not to say
25. What is the shape of a red ball? ○ Red ○ Blue ○ Square ○ Round ○ Prefer not to answer
26. Which of the following best describes your educational background or job field? ○ I have an education in, or work in, the field of computer science, computer engineering, or IT. ○ I do not have an education in, or work in, the field of computer science, computer engineering, or IT. ○ Prefer not to answer
27. Which of the following best describes your educational background or job field? ○ I have an education in or work-in/practice law or other legal services. ○ I do not have an education in or work-in/practice law or other legal services. ○ Prefer not to answer
28. What was your total household income before taxes during the past 12 months? ○ Under $15,000 ○ $15,000 to $24,999 ○ $25,000 to $34,999 ○ $35,000 to $49,999 ○ $50,000 to $74,999 ○ $75,000 to $99,999 ○ $100,000 to $149,999 ○ $150,000 or above ○ Prefer not to say

A.5 Debrief

Information on breaches your email address was part of: Thank you for completing our study. Please note that the information about data breaches we showed to you is real. Your email address, and potentially other personal information has been part of these breaches and could be used by criminals to steal your identity or access your accounts.

List of breaches your email address was part of: Below is the full list of breaches in which the email address you entered was involved according to haveibeenpwned.com. Please note that you can always obtain the same results by checking your email address on haveibeenpwned.com, which, in addition, also provides records with sensitive breaches upon the verification of your email account. Please keep in mind that this list only reflects breaches that are registered in the haveibeenpwned.com database, your information may have been exposed in other breaches.

Resources for breach recovery and further reading Here is a list of resources to help you prevent or recover from harm due your information being exposed in data breaches, as well as help you better protect yourself from data breaches in the future.

- *Resources about recovering from a data breach:*
  - Federal Trade Commission: Identity theft recovery steps
  - Federal Trade Commission: Credit Freeze FAQs
  - Firefox Monitor: What to do after a data breach
  - Norton: What to do after 5 types of data breaches
- *Resources about protecting yourself against future breaches:*
  - Firefox Monitor: How to create strong passwords
  - Firefox Monitor: Steps to protect your online identity

B Qualitative Codebook

In the following we provide our unified codebook with the primary codes, their respective counts, and their first-level sub-codes.

• **bad actors (17)**: *company sell data, hackers, department stores* • **behaviour (94)**: *continue use as before, insecure, keep using email, secure practice, email practice, insecure practice* • **cannot recall (17)**: *confused, unconcerned, surprised, concerned* • **consequence experienced (97)**: *compromised accounts, information disclosure, spam, data on the dark web, scam, attempted login, other account with same pwd, email disclosure, identity theft, social media account hacked, physical, financial disadvantage, unrecognized new account, past event, reputation, job offer missed, upset, site breached* • **consequence potentially (92)**: *spam, identity theft, compromised accounts, information misuse, financial disadvantage, scam, physical, financial account hacked, information disclosure, stalking, other account with same password, unrecognized new account* • **data not relevant (84)**: *outdated, fake data, not sensitive, unique password, not primary email, little data, will be caught by spam filter, so much data out there, account not used, unimportant password, unique username* • **data relevant (3)**: *sensitive* • **defense intended to be put into place as reaction to breach (180)**: *change password, monitor email, use secure passwords, monitor suspicious activity, monitor financial information, do not use facebook login for shopping sites, increase protective measures, change email, be more cautious, 2FA enabled, limit online disclosure, review accounts, stop using, reduced use email, check suspicious emails, signing up to websites less often, new email account, learn more about breach, reduced use site, close account, scan computer frequently, re-link security accounts, change financial information, change employer, monitor accounts, use vpn, review financial information, unique password, change username, use password manager, go after companies, learn about safeguarding, solve issues as they appear, security checkup, check financial information, protective measures, stop using email, protect email, stop using service, tor, investigate, strong password, location setting, no reuse password, be more careful, legal action* • **defense put into place as reaction to breach (226)**: *use password manager, change password, reduced use site, change emails, protective measures, 2FA enabled, change password creation strategy, unique password, no cc info in unused apps, actions caused by other breach, close account, change username, remove email from accounts, use secure passwords, review financial information, use breach monitors, be more cautious, review account information, update browser, check suspicious emails, change email, stop using site, nothing, changed info, check account, 2fa enabled, limit data disclosure, unsubscribed from mailer, change info, reviewed prior steps, monitoring, check financial, email practices, contacted company, changed email, unsubscribed, changed password, delete account, learn about breach, antivirus, called credit card company, recover hacked account, careful disclosure, no reuse password, strong password* • **defense put into place pro-actively before breach (40)**: *use secure passwords, 2FA enabled, be cautious, change password, don't answer phone calls, review financial information, unique password, use password manager, monitor accounts, monitor emails, unique email, protective measures, monitor credit reports, spam filter, stop using site, change email, account not used, monitor financial information* • **do not know hibpwnd (2)** • **feeling (929)**: *unconcerned, concerned, violated, annoyed, negative, skeptical, uncomfortable, fatigued, paranoid, cautious, hopeful, upset, scared, unsurprised, would have been contacted, overwhelmed, disappointed, unsure, reassured, don't care, curious, not worried, relief, insecure, no fear, worried, unhappy, not important enough, confused, indifferent, surprised, unsafe, ashamed, regret, informed, used to breaches, no blame on company, upset* • **first breach (1)** • **immediately informed (1)** • **impact (525)** *impact little, impact none, impact large, impact positive, impact unsure, impact negative, unconcerned* • **needs more info (1)** • **not hacked into a lot (1)** • **third party (11)**: *bad security, good security at company* • **unclear (2)**