Kejsi Take*, Kevin Gallagher, Andrea Forte, Damon McCoy, and Rachel Greenstadt

"It Feels Like Whack-a-mole": User Experiences of Data Removal from People Search Websites

Abstract: People Search Websites aggregate and publicize users' Personal Identifiable Information (PII), previously sourced from data brokers. This paper presents a qualitative study of the perceptions and experiences of 18 participants who sought information removal by hiring a removal service or requesting removal from the sites. The users we interviewed were highly motivated and had sophisticated risk perceptions. We found that they encountered obstacles during the removal process, resulting in a high cost of removal, whether they requested it themselves or hired a service. Participants perceived that the successful monetization of users PII motivates data aggregators to make the removal more difficult. Overall, self management of privacy by attempting to keep information off the internet is difficult and its' success is hard to evaluate. We provide recommendations to users, third parties, removal services and researchers aiming to improve the removal process.

Keywords: privacy, people search websites, risk perceptions, online harassment, data brokers

DOI 10.56553/popets-2022-0067
Received 2021-11-30; revised 2022-03-15; accepted 2022-03-16.

1 Introduction

In the United States, the tensions between transparency and privacy in the context of public records, collected and managed by the government, have continuously been a point of contention and research [34, 45]. Nowadays, sources of publicly available Personal Identifiable Information (PII) have moved online, been aggregated with other sources, and have become more easily accessible to both regular and malicious Internet users. While at the end of the last century opponents of abortion rights were publicizing various types of PII with the intention of inciting attacks towards abortion providers and patients [40], attackers nowadays use information available on the internet to harass their adversaries [30, 38]. For example, one of the most common types of online harassment that involves use of PII is doxing, defined as the release of someone's personal information onto the Internet with the intent to do harm [15]. Personal information available on the internet can be also used in other vectors of abuse, such as identity theft or social engineering.

Some of the most easily accessible sources of publicly available PII are People Search Websites, sites that display compiled reports of users' information. Their main source of information are data brokers, entities that specialize in aggregating users information from extensive online and offline sources. Anti-doxing guides often mention People Search Websites as one of the methods attackers might use to collect information about a target and recommend removal as a protective measure [4, 28]. Due to the prevalence of these sites, the onus to protect ones' privacy and avoid these consequences falls on the user. While there are many online resources aiming to guide internet users through the self-removal process [9, 52, 53] and newly emerged removal services [13, 16, 27, 37], the experiences of removal of information from these sites have not been studied before. Our research goal is to understand the steps required during self-removal and the experiences of using paid removal services. By gaining insights on participants' perceptions of the People Search Websites, their threat models and the experience of removing information, we can provide recommendations to (1) users who might be interested in taking these measures, (2) researchers working on tools to facilitate the removal process, (3) legislators and non-profit organizations advocating for privacy rights.

We conducted 18 semi-structured interviews with participants who have removed their information from People Search Websites. These interviews provided rich qualitative data on participants' motivations, percep-

***Corresponding Author: Kejsi Take:** New York University, E-mail: kejsitake@nyu.edu
Kevin Gallagher: DCentral/INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, E-mail: kevin.gallagher@tecnico.ulisboa.pt
Andrea Forte: Drexel University, E-mail: af468@drexel.edu
Damon McCoy: New York University, E-mail: mccoy@nyu.edu
Rachel Greenstadt: New York University, E-mail: greenstadt@nyu.edu

tions, and the removal experience. Our participants' threat models include harassment risks, physical security risks, risk to family members, financial and reputation risk, as well as risks stemming from further processing of information, such as re-identification and behavioral profiling. In addition, participants' perceive People Search Websites as enablers of online and offline abuse, facilitating access to a targets' PII and therefore causing significant personal and societal consequences.

Our participants describe complex removal experiences. Those who removed their own information report encountering multiple usability issues on these sites that they perceive as "dark patterns", such as having to provide additional documentation, requirements to create an account, pay-walls, as well as difficulties with finding removal instructions. They attribute these difficulties to the value that their data holds for the business model of the People Search Websites. Whether using a removal service or doing it themselves, participants describe that removal success is difficult to gauge and keeping information off the internet requires additional and continuous efforts.

While, these results are not entirely surprising, they provide longitudinal insights on the obstacles faced by even the most motivated users that can aid in the development of privacy enhancing technologies, transparency mechanisms, or regulatory approaches.

2 Background

As context for our study, we provide an overview of the People Search Websites, as well as users' removal options. We also discuss data brokers and outline the common sources of information, as identified in previous work.

2.1 People Search Websites

People Search Websites compile user reports by collecting information about individuals such as age, current and past addresses, connections to relatives, email addresses, online profiles, and sometimes court and civil records. The services often provide a limited amount of information for free, with the option to purchase more detailed personal information for a fee. In Figure 1, we show an example of an anonymized user report from MyLife, one of the People Search Websites, that also includes a reputation score in user reports.

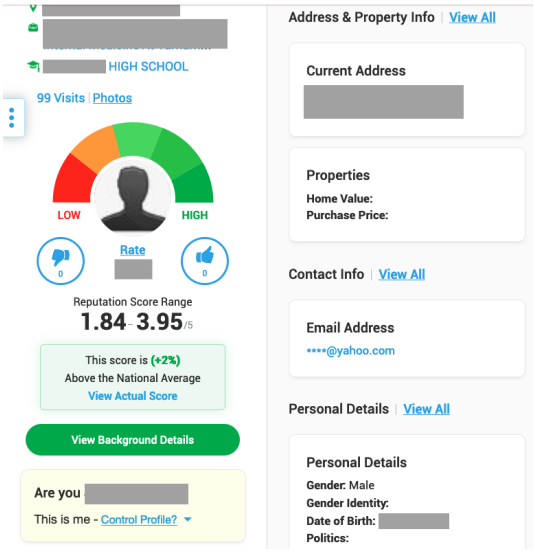

Fig. 1. Partial screenshot of an anonymized user report from mylife[.]com, one of the People Search Websites

While there is not a unified formal definition for People Search Websites, prior work defines them as one of the three main services that data brokers offer [12] or as a separate data broker type [20]. No matter which definition is correct, it is clear that they source information in a similar way to that of data brokers, whose main sources of information are: (1) government sources, (2) other publicly available sources and (3) commercial sources. Some examples of state and government data are property records, voter registration, court records and licence information [12]. Often data brokers also collect information directly from consumers or other data brokers [20]. Sometimes they collect and aggregate user information from multiple sources. For example, one of the data brokers included in a 2014 Federal Trade Commission report has about 3,000 data segments for nearly each U.S. consumer [12].

The size of the audience of these sites' is difficult to estimate. However, Spokeo, one of the largest People Search Websites, declares serving about 20 million people a month and answering 500,000 searches/day[1]. While, some of these sites offer users the option to remove their information if they wish, the removal for users not residing in California, falls into a "grey-area", as it is not included in any federal U.S. regulation. In this paper, we focus on People Search Websites, as easily accessible sources of PII, the perceived risks stemming

1 https://www[.]spokeo[.]com/about Accessed on 02.17.2022

from such increased accessibility and user information removal experiences.

2.2 Removal Resources

The proliferation of People Search Websites that publicly display one's private information has led to the emergence of businesses that offer information removal services for a fee [13, 16, 27, 37]. These services are still relatively new and little is known about how they work. The number of sites they promise to remove information from varies across these services, ranging from 37 for DeleteMe [13], to about 6,000 for Kanary [27]. The plans and prices they offer also vary and sometimes allow for the inclusion of family members.

Privacy organizations [11, 17] and independent researchers [9, 53] have compiled free guides for removing information from data brokers and People Search Websites. In addition, Privacy Bot [7] is an open-source project aiming to automate the process of sending opt-out requests. Although none of our interviewees had used it, we reviewed online articles from people who tried it [10, 21]. These articles warn about the technical skills required [21], as well as about the difficulty of fully automating all the removal steps required by the People Search Websites [10].

3 Related Work

Perceptions of the People Search Websites and removal experiences are relatively unexplored in prior work. To better understand the ecosystem of these sites and make recommendations to stakeholders, we explore the perceptions and experiences of users who sought removal. We find that our work is closest to prior work exploring user privacy rights and factors impacting security and privacy behaviours.

3.1 Privacy Rights

Information removal from People Search Websites is not federally regulated in the United States, but the CCPA gives California residents the right to request their personal data to be deleted [25]. "The Right to be Forgotten" grants internet users located in the European Union the right to delete their personal data from websites, under certain conditions [41].

In contrast to "privacy by default", users' privacy choices under both these regulations are governed following the "notice and choice" framework, giving the individual options for the collection and use of their personal information, through privacy policies or opt-out interfaces. Prior work has studied the cost associated with reading privacy policies [32] and dark patterns in opt-out interfaces [22, 23, 36, 42]. Fewer studies have looked at data deletion specifically. Habib et al. analyzed 150 English-language websites and found that while 74% offered deletion, less than 20% offered a direct link or tool, forcing the user to navigate complex options [23]. However, they found that GDPR contributed to increased deletion options.

Our research complements these studies by investigating users' removal experiences in the context of People Search Websites, showing that users encounter similar obstacles and dark patterns as prior work on data deletion and opt-out choices. Our study also brings forward longitudinal aspects of the removal experience, such as the possibility of the re-emergence of information.

3.2 Security and Privacy Behaviour

Previous work has explored factors that influence users' security and privacy behaviour. Some such factors are costs associated with protective measures [54], skill level and socioeconomic status [39], experience of privacy violation, or risk perception [24]. We extend this work, finding that perceptions of risk motivate users to remove their information from the People Search Websites. Further, our findings show that there are certain groups of the population, such as content creators and activists, that are particularly motivated to remove their information from the internet. Our findings are in line with prior work that has found that new technologies and the internet introduce new risks for marginalized or high-risk populations [29, 31, 43]. Our participants' experiences echo the findings that the internet, along with the simplified accessibility of PII available on the People Search Websites, facilitates new vectors of online and offline abuse.

However, a set of factors can prevent internet users from transforming their risk perception into protective behaviour and achieving their desired privacy levels. For example, Acquisti et al. found that privacy and security decision making is affected by psychological factors such as, but not limited to, incomplete and asymmetric information, bounded rationality, as well as eco-

nomics factors such as privacy under-supply [2]. As an additional potential factor hindering adoption of privacy behaviour, Solove describes the self-management of privacy as a vast, complex and never-ending project that does not scale [46]. Our participants' experiences suggest that removing their information from the internet is indeed a never-ending task, making privacy self-management considerably harder.

4 Method

To address our research questions, we conducted semi-structured interviews with 18 individuals between May and August 2021, lasting between 20 to 75 minutes.

Ethical Considerations. This study, including the screening survey, interview script, recruiting materials/methods, and data storage methods was approved by our Institutional Review Board (IRB). Quotes used in this paper have been carefully selected and redacted to avoid identifying participants. Throughout the paper, we avoid using participant identifiers and we use gender-neutral pronouns to ensure participants' privacy.

4.1 Participant Recruitment

Recruiting participants who had attempted to remove their information from People Search Websites posed some difficulties. We started by posting on Reddit subreddits related to privacy[2] [3]. We also advertised on other subreddits that included previous conversations about data removal. Due to the difficulty of finding users who adopt privacy behaviors, recruiting in Reddit privacy communities has been a choice in previous studies of Privacy-Enhancing Technologies [18, 33]. However, while we recruited our first participants this way, they were predominantly male and generally privacy and technology-savvy. To obtain a more diverse sample, both in gender and backgrounds, we reached out to the authors' professional contacts. We used snowball sampling [6], asking them to refer us to other participants.

When the methods mentioned above did not generate enough participants to approach saturation (the point at which interview data became redundant and yielded no new insights), we also used an active recruitment method on Twitter and Reddit, reaching out to individuals who had publicly spoken about using information removal services on social media. While prior research has studied downsides of this recruitment strategy [19], we mitigated the privacy impact of this method with an anonymized survey link (in accordance with our IRB-approved protocol), which made it impossible for us to identify the method with which the participant was recruited. This multifaceted recruitment strategy was adopted partially to overcome the difficulties of finding participants who have sought removal, as well as with the intention of including individuals with diverse risk perceptions. We also reviewed these posts to inform our analysis and to refine the interview protocol. The final interview script is included in the Appendix.

Participant Demographics. Overall, we interviewed 18 participants (nine male and nine female), with at least one representative from almost all age groups included in our screening survey (25—55+). We compensated participants with a $25 gift card for Amazon. Three declined the reward. 14 of our participants were located in the United States. Based on the interviews, we inferred that three of the other ones were dual-citizens or Americans living abroad. More specifically their experiences were similar to participants based in the U.S. and helped us achieve saturation, as gauged by repeated themes. We found that one of the participants did not have an apparent connection with the U.S. and had not removed their own information, but helped a friend do so. While our screening survey did not successfully filter for this case, we found the interview relevant and insightful and so we include their experiences in the analysis. Only one of the participants noted on the survey that they used the CCPA to accomplish removals.

4.2 Interviews and Analysis

All interviews were conducted online using virtual-conferencing software based on participants' preferences. They were audio recorded and transcribed with their permission. Transcriptions of interviews were analyzed using thematic analysis [8], an inductive, iterative approach to coding with the goal of finding themes in the data. The analysis was performed by the researcher that led the interviews, to ensure an uniform analysis of the data. However, to avoid subjective bias, after the first five interviews were conducted, initial themes were discussed with one of the other authors.

2 https://www[.]reddit[.]com/r/privacy/

3 https://www[.]reddit[.]com/r/privacytoolsIO/

At this point, we made small adjustments to the interview protocol to account for the emerging themes. After 14 interviews, the themes were discussed among the authors and we determined that, although in some areas the analysis approached saturation, we lacked perspectives on the removal services and concentrated further recruiting efforts there. After 18 interviews were conducted, the first author continued iterative rounds of coding on the full dataset to refine themes and discussed and condensed them with the second author.

5 Results

We present our results focusing on these areas: risk perceptions, perceptions of People Search Websites, removal experiences and removal outcomes.

5.1 Risk Perceptions

We examined the perceived risks stemming from publicly available information. Participants conceptualized risks related to the direct use of their data by unauthorized third parties, such as online harassers or identity theft criminals. They also expressed concern about indirect use of PII, through behavioral profiling and re-identification.

5.1.1 Direct Use of Publicly Available PII

Most of the concerns related to the direct use of PII were targeted online harassment and financial concerns due to identity theft.

Harassment Risks. Fourteen of the participants mentioned at least one form of harassment as a perceived consequence of publicly accessible PII. The most common types of harassment concerns mentioned are doxing and stalking.

Six of the participants had experienced targeted online harassment, such as doxing, online trolling, bullying and impersonation, as defined in prior work [48]. For four of them the harassment experience was the motivation for seeking removal and the other two participants had removed their data preemptively, in anticipation of being targeted. Three of the six participants who had been harassed, perceived that they had been targets of ideological harassment because of their activism, including publicly raising privacy issues online or expressing political beliefs. Harassment campaigns as a self-preservation mechanism by attackers or as a result of ideological differences between targets and attackers have been discussed in previous work [14]. Seven participants who were involved in content creation for public consumption, either as a hobby or as part of their job, perceived that they faced increased risk due to either the choice of subjects, or the specific way they might be portrayed. Three other participants had not experienced online harassment themselves, but knew someone in their close circle who had.

Mental Health Impact. Two participants were particularly concerned with mental health consequences of online harassment. One participant, an occasional social media user, described a feeling of self-blame instilled by the fear that information shared online could be weaponized by harassers: *"I feel so stupid for putting anything online that suggests I'm here in [city], [...] [with online harassment] there can be a lot of like victim blaming, you know, blaming myself, as like the victim of this harasser and I felt like a shame that it was happening. I question like what have I done wrong, or you know why this is happening to me?"* Such feelings of self-blame can lead to self-censorship, or the restriction of participation in online discourse, a practice that four of our participants mentioned following at various degrees in fear of facing retaliation.

Another participant suggested that women bear the heavier weight of online harassment: *"The Internet is a more potentially dangerous place for women generally and my guess is that men just don't think about these things"*.

Reputation Risk. Three participants perceived reputational damage as a risk of having their private information easily accessible. For example, one participant mentioned risk to their reputation as a result of content leakage [48], a type of online harassment that could result in reputation damage: *"...they [the attackers] could put spyware on my computer or maybe they could discover something embarrassing about my private life or my sexual life or something like that and um post that on Twitter."* While information in the People Search websites might not be enough to result in a hacking campaign, it can be integrated in phishing or social engineering attempts.

Two participants mentioned the alleged court records that the sites display as a potential source of reputation harm. For example, one participant expressed that while they had not experienced it themselves, they worried that reputation harm could escalate and damage their social relationships: *"Someone could decide not*

to want to have a social relationship with you because they found this information [alleged court records]." We explore users' perceptions of information found in the People Search Websites, in section 5.2.1.

Risk to Family Members. Six participants mentioned being worried that their information on these sites was also linked to family members. For example, one participant explained: *"When I saw that [childhood address] and saw like all the family members, I was like well and like maybe I have made a decision for myself and I'm comfortable with having some sort of online presence, but like I don't want to drag everyone else into that. I don't want someone to be able to, like, send my mom a threatening letter."* Participants concerns illustrate that online harassment can often extend beyond the main target, to include their close circle of family or friends.

Physical Safety. Participants discussed how the lack of privacy might contribute to the materialization of physical safety threats from an online harasser. For many participants, hiding their physical address was a primary motivation for seeking removal from the People Search Websites. For example, one participant says: *"He [the attacker] could easily find out where I live. Does he live in the same state as me, does he, you know, is he gonna send mail to my home?"*

Five participants were particularly concerned with physical security threats, caused by the availability of addresses on these sites to online stalkers, e.g. *"Stalkers nowadays have been mostly online, but with the availability of people's addresses I think it can escalate beyond online."*

Financial Risks. Eight of the participants mentioned identity theft as one of the reasons they sought data removal. However, only four of them mentioned taking other identity theft protective measures, such as having set in place credit freezes or using an identity theft monitoring service. None of the participants had experienced identity-theft, but two of the participants knew a friend or a family member who had and associated this event with becoming more conscious about how publicly available information could be used.

5.1.2 Indirect Use of Publicly Available PII

Risks described in the previous sections were a direct result of the use of PII by individual or groups of actors to cause harm to a target, often as a consequence of something they did or said. In contrast, risks described in this section stem from the trends and data points that can be inferred from information that users give away.

Re-Identification. Three participants mentioned worries related to the linkability among different types of PII. Two were particularly concerned with the limited number of data points necessary to uniquely identify unique U.S. citizens e.g.: *"You can uniquely identify Americans using their information, their zip code numbers or like phone numbers and addresses and names and stuff."* The feasibility of a large-scale de-anonymization of the U.S. population has been discussed in previous work [47]. To reduce this perceived risk, these three participants limited and obfuscated the amount of PII they gave away, by creating alternate identities, as well as giving websites or services different information.

Behavioural Profiling and Surveillance. While eight participants discussed being concerned with online tracking and surveillance, three of them mentioned the Cambridge Analytica incident [26] as the event that made them re-think the data they had given away in the past and look into ways to minimize them going forward. While they did not use the word profiling, one participant explained that this incident provided them with insight that additional clues can be inferred from what initially looks like superficial information: *"I would also say it's a bit more cloaked personal habits, or preferences, like if I shop for certain things you know having read about how Christopher Wiley, like Cambridge Analytica was really kind of pulling data and was able to construct patterns on people based on their preferences.[...] I would say like the consequences being that personal information is a lot more telling than you know what I get for dinner."* For most of these participants, removal from People Search Websites was just one of the many measures taken to reduce their digital fingerprints, such as minimization of social media usage, compartmentalization of information given to online services, or even requesting data deletion when possible.

In addition, two other participants expressed concern about the impacts that profiling can have, especially for internet users who are members of marginalized groups, e.g. *"I know that there are additional uses of that information, like you know, I mentioned like insurance company, maybe employers things like that, you know, contributes to like an overall um environment where people who are vulnerable are made, maybe even more vulnerable and then their ability to, you know, be free online is affected."*

5.2 Perceptions of People Search Websites

We find that participants' perceptions of the People Search Websites are connected to the business models of the sites. Participants expressed concern about the handling of their information and the potential the information on these sites has to be used by attackers. While they had some suspicions about the ways the data is collected, they also described a sense of helplessness associated with fully understanding these practices.

5.2.1 Dubious Information Handling Practices

For four of the participants, negative perceptions of the sites stemmed from the way they perceive their information is handled and used to increase profit. Participants were particularly concerned with two aspects, the "distortion" of the information to increase traffic to their websites and the dissemination of information to other parties. Two of the participants described thinking that the People Search Websites added artificial information to lure users into paying to access the full reports.

One participant described the addition of untrue datapoints, along some real information, a practice they thought aimed to gain the trust of website visitors: *"It seems like they were adding information that I don't understand how they would have even gotten it together, so they were adding things like marital status, number of kids and that sort of thing, [...] It was like they had enough data for people to think it was accurate you know many of my age but then they were adding more data to try and get people to buy reports."*

In addition, another participant discussed a similar observation, the distortion of existing information to increase and incite curiosity in website visitors: *"You could have an incident, where you just have like a minor traffic violation and all of a sudden when you Google your name it says so and so may have arrest records. [...] The people services rank the parking fine on the same level as you know you were convicted of assault, with a deadly weapon."* Similarly, two other participants described the suspicion that the People Search websites are selling their information to third parties, who later use it for marketing purposes. For one participant this perception materialized when they found out that the information on one of the People Search Websites coincided with the address included in some junk mail.

5.2.2 Perceived Information Sources

Overall, most participants agreed that these websites gather information from a variety of sources. Most participants mentioned publicly available information and government sources as a source of information for the People Search Websites. Seven of the participants mentioned voter rolls and five of them mentioned property records. To reduce the risks of having their information publicly available, two of the participants attempted to remove their name from the voting records. However, they found that the only way to do so would be by giving up the right to vote, a trade-off that the participants explain that they were not willing to make. One of them said: *"It was very difficult to have that important part of being a citizen, contingent on giving up that part of my privacy."* Three participants also mentioned arrest or court records as a source of information. One of the participants expressed concerns related to the common perception that "nothing dies in the internet", making such public records an unnecessary extension of the punishment that an individual received: *"It's really sad to me when I'm looking for someone and what comes up on immediately is mugshots from a county website that are just never going to be gone and it's like that that's a punishment that lasts well beyond, perhaps someone sentence, or after they pay a fee for a misdemeanor."*

In addition, participants perceived that the information these companies have is used to extrapolate additional data points. For example, two participants described their mental model of an automated aggregation process, discovering family relations based on common addresses e.g. *"Uh yeah there's some kind of program that's pulling it together to go, you know, person A and person B shared an address and based on their dates of birth, which are public, these people were probably you know, a parent and child."*

Social media is another source of information frequently mentioned by participants. However, one participant described the suspicion that the main purpose of the data online social networks collect is mainly used for targeted advertising: *"I mean I think the social media stuff is ah not more anonymized I guess it's more like you can target [name] because [they are] between 30 and 35 and [they live] in an urban center and [they like] this kind of clothing so you'll get an ad like that. I'm not sure, I mean I think they're packaging and selling information I don't think that Instagram is like here's [their] home address but maybe they are."* The fact that Facebook has used data collected from data brokers for their targeted advertisement ecosystem has been well-

documented [51]. Researchers have also explored attacks that could result in inferred PII by exploiting their custom audience advertising feature [50]. Six participants mentioned being unsure where the information comes from, even if they list a couple of guesses. One of the participants explained having read what People Search Websites themselves describe as potential sources, but having found these explanations vague: *"I don't have a good sense of where all this information comes from, other than you know, they talk about public records all the time, and lord knows what that actually means."*

5.2.3 Potential of Abuse by Attackers

Another negative perception of the People Search Website stemmed from personal experience with being a target of online harassment. One of the participants who had previously been harassed by a targeted online hate campaign mentioned that harassers might make use of the premium paid services that the People Search Websites offer: *"I don't know like how common it is for like 8Chan people to buy those accounts but I imagine that they probably they probably buy those."* While it is not clear where the attackers get their information, the frequency of online harassment and toxicity of content in online imageboard forums, often containing targets' PII has been well-documented in prior work [3, 44].

Low Effort Required. Another participant with a similar experience discussed the facilitated access to PII that these sites offer. Having used these services to find people as part of their journalistic investigations, they described the low efforts required to use them, while considering the possibility that an harasser might use these sites to find information about them, as a target, instead: *"I find it helpful in my work when I look someone up and there's not much about them, but I'm like 'Oh well their mom lives right here and here's her email, I could email her and be like hello I'm trying to get in touch with your daughter, and I wondered if you might be able to pass along my requests', but being on the receiving end, when it comes to harassment is like oh, there's me [name] and then there's [brother's name, father's name] and you know my brothers, and my dad and my mom."*

Low Cost of Usage. Apart from the ease of use, the same participant also expressed concern about the low cost of finding information in these sites, which also lowers the marginal costs the attackers face when using them to cause harm: *"What responsibility do these data companies have when someone gets a letter at their home because a harasser found their address in five minutes by paying $3.99?"* In summary, participants perceived that availability in the People Search Websites makes the information easier to access by attackers, therefore lowering the costs of harassment campaigns. Earlier work warns about privacy consequences of the increased accessibility as a fundamental difference between public records that are accessible only in the various localities they are kept, such as town or counties and their digitized versions available on the internet [45].

5.2.4 Overall Helplessness

Three of the participants described being not surprised that their information was available on the People Search Websites. One of them associated this attitude to a feeling of resignation caused by the vast amount of data that is related to our public online activities e.g. *"This is part of life, part of being on the Internet."*

For three other participants the feeling of surprise was connected to the level of sophistication these sites have, e.g. *"And they like to have my actual name and my childhood address, like that's the part that always feels like wow they really put a lot of effort into this and they have some deep data and its not just what I have publicly available on Facebook or something."*

5.3 Self-Removal

In this section we discuss the experiences of the twelve participants who followed a manual process to remove their information.

5.3.1 Usability Issues

Participants who attempted to do the removal themselves reported various obstacles. However, most of our participants seemed to be highly motivated and were not discouraged from taking action and at times also found workarounds.

Four participants mentioned that a small number of sites requested ID verification before they removed the information. However, two of them found ways to bypass the ID verification step, by blurring as many elements as they could or by contacting a representative and using an alleged past incident of identity theft as an excuse to avoid this requirement. While it is not clear what percentage of the People Search Websites requires

additional ID verification, it might be the case that ID verification is required in the cases where the user appears anonymous in their correspondence with the site. For example, two participants that had to provide ID, mention that they frequently use an anonymous email through an email forwarding service.

The steps required to remove one's data can be more difficult to bypass than just the ID verification. Three participants mentioned having to create an account or purchase a subscription to claim their profile and ask for removal. One explained that this process can be very difficult for people like themselves who have undergone a name change: *"The companies, the sites themselves often they won't let me do it, because I will claim one account called [first, last] my name, my current name and then they won't let me claim another account, because that's not my current name."* The participant later added that while there might be procedures for providing information about their legal name change, they felt uncomfortable sharing that information. This practice disproportionately affects some people more than others, for example, those who change their name to better fit their gender identity.

Six participants reported difficulties finding removal instructions on the People Search Websites. Four participants mentioned trying multiple times to remove their information after having received no indicators that it worked the first time. One participant described unsuccessfully escalating removal requests to higher levels: *"One of them is validnumber[.]com, I'm still in there I can't get my information removed. I've contacted them seven times so far—over seven times—and since March about two or three times a month and they just ignore all your requests. I've gone up the chain of command and went to their host of their website and they said they can't do anything."*

Another participant observed that the lack of confirmation or acknowledgement after completing a removal form requires a follow up. Additionally, three participants explained it is very difficult to know if a profile was removed when user reports are hidden behind paywalls. Other "dark patterns" reported were obfuscated or inactive removal form links, which participants perceive as a deliberate attempt to discourage users from removal and maximize profit. Dark patterns that have the potential of discouraging users from taking privacy protective options have been extensively studied in previous work [1, 22, 23].

5.3.2 Information as a Key Resource

Another common perception among participants was the belief that these websites make the removal process difficult because their main goal is to maximize profit. One of the five participants who held this belief explained non-promptness as a strategy used to avoid action demanded by users, e.g.: *"Some [People Search Websites] took multiple attempts to get it [the information] removed, [...] I requested this and you said you would remove it within 24 hours and you didn't, so it's more having to ride their ass so to speak because they just they just don't have a motivation to remove them."* The participant also attributed measures aiming to make removal difficult to the business model of the People Search Websites, e.g. *"They don't make it easy to take your stuff off, just because they are making money by selling this information."* As we explain in section 2, these sites monetize PII by selling user reports and trading information with other data brokers.

5.3.3 Empathetic Customer Service

In contrast to the experiences described above, two participants described representatives of People Search Websites as responsive and understanding when data removal was important to the individual's personal or financial safety. One participant who helped a friend facing an online harassment campaign explained that she was surprised how fast one of the People Search Websites removed the information once they were given a reason: *"We approached it from the angle [the removal], saying this is like a vulnerable adult, [...], highlighting the harassment and the danger that could follow from that and that was pretty much enough to like have them do it without asking any more questions."* Another participant also described a similar experience of receiving a positive response, after alleging that that they had been a victim of identity theft, a strategy they used to limit giving away additional information.

5.3.4 Evolution of the Removal Process

An adjacent positive observation stemmed from the long experience that one of the participants had with removing their information. They described that in the past, the removal process required mailing physical copies of one's documents, but it got incrementally easier: *"I think it's a little bit easier now than it was back in the*

day, I very vividly remembered in, um, the mid 2000s, 2006, 2007 having to physically mail, like a photocopy of my drivers license and like a letter that says please remove, you know, please remove my name..."

5.4 Removal Services

In this section we discuss the experiences of six participants who used a removal service to remove information from the People Search Websites. Three of the participants used DeleteMe [13], two used Privacy Duck [16] and one used a service targeted to current and past law enforcement officers. All of the services promise to remove information from the People Search Websites only and not other sources.

5.4.1 Understanding of Services Offered

Overall, participants have a good understanding of how the removal services work. Three of them mentioned the removal services websites as source of information.

Mostly Complete Understanding. While one of the participants did not explicitly mention looking at the website, they seemed to have an overall understanding of the process: *"I assume there I'm paying them to do is to try to go to those websites that have scraped my data from elsewhere and say I'm requesting on behalf of [full name] that you remove my data from this website.*

The same participant explained that users' privacy rights allow for the removal of the information: *"I guess you probably have a right to make this request for your private data to be removed and the company probably has to oblige."* Although removal services websites do not specify geographic restrictions, in the U.S., this is a legal requirement only in the state of California.

Perceived as Harassment Preventing Tools. One participant who had used the removal service based on a recommendation from colleagues, mentioned not being familiar with the People Search Websites and perceiving the service as an anti-harassment tool: *"My understanding was that if I signed up for Privacy Duck, it would be a little bit harder for someone who had a bad intention towards me to access my home address and other personal information and again in the case of swatting, [...] they can you know, look up your address, call the cops to say that you're a pedophile who has kids in the basement there and then people kick your door with a sniper rifle."*

An Attempt to "Buy" Privacy. Another participant perceived the removal services as an attempt to buy privacy, describing them as a way of *"throwing money at the problem"*, while others mentioned saving time as the main motivation to use these services. As discussed in section 5.3.1, manual removal is cumbersome and time-consuming, making hiring a removal services very enticing.

5.4.2 Uneasiness with Giving Away Information

All participants who used a service reported having to supply the company with their PII, that included current and past addresses, phone numbers, names and sometimes information about family members. In addition, a participant who had used a removal service targeted to law enforcement professionals, reported that in addition to their PII, they also had to provide proof of working in the field.

Two of the participants described discomfort with providing information in exchange for removal.

Privacy and Security Concerns. For one of the participants, this feeling was connected to the perception that the removal service could be hacked: *"Then they're like please give us as much information as you can, like tell us past addresses you've lived at, tell us who your family members are, tell us which phone numbers you've used and I was filling that out, I was like well I hope DeleteMe doesn't get hacked because here's all my information again, here it goes."* This quote highlights a tension between companies whose goal is to manage users' privacy and their personal data collection practices. Similarly to the self-removal process discussed in Section 5.3, to achieve a desired state of privacy, the user has to first sacrifice some of it, by supplying the removal services with additional personal information.

Uncertainty Associated with Removal Services. Another participant mentioned regretting having given information to the removal service they used. The feelings of regret were instilled by the non-responsiveness of the service and the lack of regulation of the removal services: *"I still feel very uncomfortable about that [giving away PII], especially given that they're not responsive to any of my emails [..], they have all of my data now, which is very scary given that, they're like, not a well-established or well-regulated industry."*

Participants who seek removal of their information perceive their information as sensitive and valuable. Therefore, it is not surprising that some of them found it very uncomfortable to share it with the removal ser-

vices. Even if the removal services were trustworthy and effective, these companies could still be subject to data breaches or other incidents resulting in unauthorized access. Additionally, once given away, one loses control of their information. If the company goes off-market or stops being responsive, there are no rules regulating their behaviour and therefore what happens to the consumers data is currently unknown.

5.4.3 (Dis)Satisfaction with Removal Services

Participants described various degrees of satisfaction with removal services. Three participants mentioned feeling safer after using the service. One of them renewed the service for a second year and another intended to do so in the future.

Partial Peace of Mind. However, one participant described a more neutral attitude, explaining that while they could not evaluate how effective DeleteMe is, paying for it was the best they could do, since paying for a professional expert team to remove all traces would not be affordable. Additionally, the participant saw value in feeling that at least some degree of protection will be provided: *"[I am] hoping that at least it does something or makes me not impossible to find, but puts an added barrier that might make someone who's like trying to harass me online, be like 'OK I can't find out where she lives in [city], I'm gonna send this threatening email but I'm not gonna be like I know you live on this street', you know, which may be good for my peace of mind."*

Potentially Incomplete Removals. In contrast, another participant who used the the removal service DeleteMe, was not satisfied with the result. The participant searched their name during the interview and found themselves listed on some People Search Websites. While it is unclear if DeleteMe offers removal from the particular sites the participant encountered during the search, this experience raises questions about the thoroughness of the removal services.

Lack of Regular Reports. Another participant, who used Privacy Duck, described receiving an initial report with the results found, but no further updates about the removal process or monitoring. The participant mentions receiving no response after reaching out to the service multiple times to inquire about the progress of their information removal. This experience left the participant feeling hopeless; not only due to the information they had given away, as mentioned in subsection 5.4.2, but also due to the lack of confirmation that any information had been removed: *"I think I came away feeling like it's a bit of a Wild West and I need to be, the onus is on me to be vigilant about privacy..."* Neither of the two participants who had used Privacy Duck recalled receiving removal updates.

5.4.4 Opacity of the Removal Services

Participants who used DeleteMe discussed regular reports as indicators of receiving the service they are paying for. However, one of the participants found the reports opaque.

Necessity for Further Contextualization. They explained that DeleteMe reports lacked context and made it difficult to understand if the removal was decreasing their risk exposure. For example, while the report mentions removing their information from a number of websites, the participant was not sure if those were the most popular People Search Websites or the removal was just *"a drop in the bucket"*.

They described a way to contextualize the removal that would make it easier to understand e.g.: *"Someone who's looking for you is now going to have a much more difficult time, or someone who might have been able to find your address in two minutes, now would have to spend 25 minutes and are they going to stop at the 20 minute mark?"* However, they recognized that it might not be in the removal service's best interest to reveal the scope of the problem: *"I don't think you could really trust DeleteMe to be like well we got rid of 31, but like they're really 3000 so why are you paying us, it's completely useless"*

Lack of Transparency. Another participant described that DeleteMe reports contain checklists of sites removed and the types of information (i.e. address, telephone number) removed from those sites, but the actual PII they found is not included in the reports. While this might be a choice aiming to protect users' privacy, in the cases when a user has a common name, it is difficult to tell if the service is removing their data or that of other people with the same name.

5.5 Removal Outcomes

As mentioned above, our study aimed to understand the experiences of two groups of participants, those who went through the removal themselves and those who used a service. Both groups of participants were unsure if the removal measures they took are final and explained that various continuous measures are required.

5.5.1 Removal Is Not Final

Eight participants expressed doubt that the removal might not be the final step towards guaranteeing their desired level of privacy.

Perpetual Information Sources. For three participants, the suspicion that their information is not fully removed from these websites was related to the People Search Websites' business model, which includes perpetual collection of data from different sources e.g." *"Um I kinda have my doubts. I know they say that it's gone, but I can't really prove that it is. There's a good, there's a possibility that there are still data sources that have all of my information on it, they are just not publicly displaying it and so I think there's definitely a possibility that that data could still end up transferred to someone else and then that person could publicly display the information, so I don't. I only feel a little bit better about the removal process."* Another participant described a similar feeling of surrender. They perceived that due to the speed the data is collected, more information might have been collected on them from other sources while they were going through the deletion process.

Continuous Effort. Similarly, one participant who used a removal service described a conflict between the need to use online goods and services and the information they have to supply in order to so. They explained that compounding effect makes the attempts to be private more difficult: *"It just feels like a never ending, impossible battle to try to keep private information off the Internet and that really sucks."* Another participant who used a removal service viewed the abundant data sources as a reason they need to continuously pay for the service: *"My understanding of the services is that this is something that you could pay for perpetually,[...] because the other thing [that happens] is they sort of feed off each other or they grab the information from other data brokers, who are doing the same thing, so I don't expect ever to get a report that's just 100% like you're in the green and you could cancel, you know, like I think as long as I'm concerned about my information popping up online as recently I would just continue to pay for it indefinitely."*

Similarly, someone else described the removal as *"band-aids on the problem"*, explaining that they did not see the removal as a solution. The participant expressed that the problem shouldn't exist in the first place and the personal safety of internet users shouldn't be in the hands of these companies.

The participants' suspicions are not unfounded, given that data brokers gather information from a variety of sources and continuously buy and sell information to each other. Additionally, a 2014 FTC report found that one of the data brokers being studied obtained consumers' contact information from twenty different sources [12]. Given that the entries that users are trying to remove could be owned by another data broker who sold it to the People Search Website in the first place, a never-ending cycle of sale of personal information is created.

Reappearance of Removed Information. Suspicions that the removal is not final are verified by experiences of three participants, who noticed that data reappeared after some time. Three of the participants mentioned having recently completed the removal and had not monitored the websites yet, but described the intent to do so because of anecdotal evidence of information re-appearance. One of them described the difficult and continuous nature of privacy self-management: *"Privacy isn't a sprint, it's a marathon, you always have to be working at it, putting in the time."* One participant who had been engaged in the removal process for 20 years, the longest in our sample, confirmed that continuous efforts were necessary: *"It feels like whack-a-mole, like you keep trying to, you know, you suppress information from one site and then it pops up on another site and then it pops up on the same site, you know, you know six months or a year later, so it feels very unsatisfying because there's no real way to get this done that sticks, after I've been doing it for almost 20 years now."*

Meanwhile, four other participants monitored websites manually, or set up Google alerts for their name and did not observe data re-appearance. It is difficult to infer from our interviews why the data seems to have re-appeared for some participants and not for others. However, two of these participants mentioned having taken other measures to hide their address, such as using a virtual or Post Office (PO) mailing box.

Similarly, the efficacy of removal services is difficult to evaluate. Five participants who used the removal services mention not taking any monitoring measures, and thus being unsure if the information had reemerged.

In general, the predatory business model of People Search Websites works in the favor of the paid removal services. As long as these websites continue to gather information from other sources, internet users who want to conceal their information will have to either spend time on removal themselves or use a paid service.

6 Discussion

Our results provide insights on the removal process from People Search Websites. In this section, we discuss potential limitations and summarize the key insights.

6.1 Limitations

We did not aim to limit the study to the U.S. and also interviewed participants from Europe and North America (non U.S.), who filled our initial survey and were open to participating in our study. However, we discovered their removal experiences were related to People Search Websites containing their U.S.-based information. More research is needed to understand the scale and impact of equivalent sites in other countries.

Our participants are not representative of the general U.S. population and their privacy concerns and experiences likely are not either. For example, participants we interviewed have high levels of education. Some of the recruitment methods we used have disadvantages outlined in previous work [6, 19], but the recruitment difficulties we encountered hint at a low rate of internet users investing in information removal making it essential to use a variety of recruitment strategies.

We were surprised at how few participants reported removing information reactively as a result of experiencing online harassment. It might be the case that such participants avoid the platforms where we advertised our study or are not willing to talk about their experiences, due to anonymity concerns.

Secondly, our study is of an exploratory nature, where we aim to understand the motivations behind the removal and the removal experiences. We make no effort to draw statistically significant conclusions. Further, even though we used multiple methods of recruiting, advertising in privacy specific subreddits might have introduced bias in our data. Finally, given that our analysis is based on self-reported experiences, it is possible that that participants over-claimed or omitted their privacy perceptions or behaviours.

6.2 Key Insights

Our results provide insights on the wide array of risk perceptions that motivates removal, the difficulty of privacy self-management due to information asymmetry and under-supply of privacy and measures that participants found successful.

6.2.1 Wide Array of Perceived Consequences

Participants' perceptions of privacy invasion and its consequences sheds light on the consequences that publicly available information can have on users, both through direct and indirect usage. At the individual level, our participants are worried that easily accessible PII might lead to harassment, threats to their own personal safety or that of their families, as well as mental health and financial consequences. At the collective level, one of the participants expressed concern how these services may disproportionately impact marginalized communities through the their information for behavioural profiling. Additionally, participants' experiences show that privacy is particularly important for writers, journalists, activists and other content creators, who might face backlash for their activities and opinions [14]. While some personal data has been historically available to the public, the proliferation of People Search Websites makes these data more accessible to attackers. Increased accessibility also decreases the cost of harassment campaigns or identity theft attempts.

6.2.2 Information Asymmetry

A common theme across participants' discussion of data sources, business models, and removal outcomes, is the lack of understanding how the People Search Websites and the removal services work. This finding echos prior work discussing information asymmetries between internet users and data holders [2]. Defined as users' unawareness about how their information is collected, disseminated, and used, information asymmetry is a reason why consumers cannot achieve desired levels of privacy. Not fully understanding the data collection practices of People Search Websites makes it difficult for internet users to protect their information from ending up on these sites. As an additional source of risk, users are uncertain what happens to the data they have to supply to the People Search Websites or removal services.

6.2.3 Questionable Business Practices

Prior research has studied the under supply of privacy, caused by the increasing value (and success) of monetizing personal data [2]. Participants perceive the People Search Websites and their questionable business practices as drivers of such monetization. Their experiences suggest that these sites might deliberately make the

removal process difficult to navigate because the consumers' data is essential to their business model. Similarly, the reemergence of information and the presence of artificial, almost click-bait data, could also be due to their attempts of maximizing profit by incetivizing website visitors to buy subscriptions or full reports.

Due to their lack of transparency, participants perceive removal services as equally dubious. The lack of contextualization of how effective the removal attempts are was a repeated concern. While some of the removal services provide regular reports explaining the number of removals completed, participants found them unsatisfying and incomplete; we provide recommendations for addressing this concern. Our interpretation of participants' experiences is that the difficulty of the removal process is monetized by removal services. While further research is needed to understand how they work, it appears like removal services have an almost symbiotic relationship with the People Search Websites. As long as the removal process is time-consuming and difficult to navigate, removal services can attract customers.

6.2.4 Privacy Self-Management

Many participants in the study were highly motivated to reduce their digital footprint using self-managed personal information removal techniques. They encountered various obstacles during the removal process, such as dark patterns and requests for additional information. Participants perceived that more data is continuously being collected about them, a suspicion that was verified by the experiences of re-emerging information. Therefore, keeping personal information off of the internet requires the continuous allocation of resources. Whether those resources are time for repeated manual removal or money to keep hiring removal services, the cost of privacy protection keeps increasing. Some participants went further and adopted additional practices to conceal their information, by attempting to conceal their physical address through the use of a Post Office(PO) box and by limiting the information they give away, further increasing the marginal cost of acquiring protection.

These findings support one of the possible explanations of the so called "privacy paradox" [35] – the one that attributes the reason for people's failure to manage privacy effectively to the fact that managing one's privacy is becoming increasingly harder [46]. Our findings imply that the People Search Websites make it difficult for users to achieve a desirable state of privacy, due to continuous amounts of time and money required for the removal making privacy-self-management very difficult and only accessible to internet users who have an abundance of such resources. The other side of the discussion centered around this paradox, suggests that people do not adopt privacy behaviour because they ascribe a fairly low or non existing value to their privacy [5], a perception that does not seem to be true for our users.

6.2.5 A Step in the Right Direction

Even though very difficult to accomplish, removal seems to be successful towards providing some degree of protection. Besides providing a degree of peace of mind to some participants who go through the removal process, one participant attributed the early interruption of a doxing campaign, they were a target of, to the proactive data removal they overtook. While we can't know for sure why the attackers gave up earlier, the participant believed that their information was more difficult to find and the harassers were less incentivized to keep looking.

7 Recommendations

Our qualitative analysis of perceptions and removal experiences from People Search Websites is a first step towards understanding how these sites monetize users' PII. We use several key insights from our analysis to inform recommendations for stakeholders to improve the removal process.

7.1 Users

When discussing manual removal, multiple participants mentioned following online guides and instructions, and emphasize the need to make detailed notes about each step of the removal process. Participants described repeatedly demanding removal, before their request was granted. These experiences demonstrate that much labor and attention may be required if manual removal is to succeed. Keeping track of each step could be helpful in following up.

Additionally, we find that removal services, even though achieving questionable results, could be a way to pay for privacy by saving one's time. While we cannot make any recommendations on which of them is better

than the others, we suggest that they are used along with continuous monitoring and manual removal in the cases when the removal services fail to remove information from some sites.

7.2 Removal Services

Our findings have implications for the removal services themselves. We show that users demand more transparency, particularly in regards to how they store, process and use user data. Additionally, regular reports, for those who offer them, seem to give users confidence that these removal services are at least somewhat effective. We also recommend that removal services contextualize removal reports to help users make sense of them, for example by calculating the number of search results relevant to the user before and after the removal.

7.3 Researchers

We invite researchers to work on solutions to make the manual removal process easier. While there have been attempts at automated methods [7], it is not known how well they work. For instance, Privacy Bot sends bulk automated emails to the sites, but it is not clear how effective this is, especially in cases when websites requests form completion, a requirement reported by some of the participants. User experience studies could look into how such automated removal scripts compare to removal services. Moreover, it is unclear whether and how many People Search Websites exist and monetize publicly available information. It will be important to measure the scale of this phenomenon and understand if and how these services are interconnected. It would be particularly helpful to be able to enhance transparency of information flows between these websites themselves, as well as third parties.

Estimating the scale of the problem could also help researchers come up with instructions to help people with different backgrounds. resources, and technical skillsets to remove their information from the internet. Additionally, researchers or other non-profit organizations could offer detailed guides, automated scripts, or consulting help particularly tailored to high-risk groups.

7.4 Other Third Parties

In this section we list our recommendations to third parties, such as regulators and non-profit organizations.

We invite regulators to investigate dark patterns in the removal process of personal information from the People Search Websites, as well as to take regulatory steps towards prohibiting the use of information supplied during the removal for any other purpose. We also suggest they re-evaluate the protected categories of citizens that can opt out of their information being included in public records. For example, some states allow removal for domestic violence targets and some professionals in law enforcement or medical staff [49], but these lists could be expanded to also include journalists, researchers, and activists.

Experiences of our participants indicate that the removal processes are continuously changing; correspondingly, we recommend continuously updating removal guides as an essential service to groups of individuals whose personal safety might be threatened. Additionally, we suggest the creation of an organization that overlooks and audits the People Search Websites ecosystem. This organization could lobby those with poor opt-out processes to improve them, as well as detect poor or unethical practices, such as the regeneration of removed information. These findings could result in a scoring system to help users understand which sites to trust, and also fines if they bring information back after deleting it.

8 Future Work

To the best of our knowledge we are the first to qualitatively analyze users' motivations when seeking PII removal from the internet, as well as the users perceptions of the People Search Websites. In section 7 we outline recommendations for users interested in removal, third parties, removal services and researchers, further work could be focused in implementing and evaluating these solutions. Due to the qualitative nature of our study and the limitations of our recruiting efforts, these findings are drawn from a small sample. To validate these findings and extrapolate generalizable conclusions, a larger study would need to be conducted. Additionally, it is not clear how many people know about these sites existence or that it is possible to request removal. Further research could focus on providing some more clarity in regards to this question and study the reasons that in-

fluence internet users decision in deciding to or not to take removal action.

9 Conclusions

Publicly available Personal Identifiable Information (PII) on the internet can be misused for many purposes from identity theft to online harassment. Removing such information is an important defense against such threats for all internet users and particularly researchers, journalists and activists who might be targeted due to their activities. Our results show that participants believe People Search Websites decrease the costs of such attacks by facilitating access to personal and potentially sensitive information.

By studying users' experiences with personal information removal from the internet, we find that the manual removal process is time-consuming and involves navigating a series of obstacles and dark patterns. Yet, while participants perceive the removal service as a way to "buy privacy" and save time, we show that the removal services are not thorough, fully transparent, or even responsive. These difficulties, as well as the lack of understanding about how People Search Websites and removal services work highlight the information—and by extension power—asymmetry between our participants and these companies.

Participants perceive that the business model of People Search Websites is very dependent on the personal information of internet users. Additionally, the removal service depend on the existence of these sites. This chain of dependence on the monetization of user data leads to privacy under-supply.

These two factors, information asymmetry, and privacy under-supply make it difficult for even committed users to keep their data off the internet. This finding supports explanations of the privacy paradox that suggest self-management of privacy is impeded by psychological and especially economic factors.

While the monetization of users' personal information is essential to the business model of data brokers and unlikely to change anytime soon, we provide suggestions to help prevent consequences related to its abuse. Our recommendations for researchers and third parties aim to make the removal process accessible to users of all skills and backgrounds. Recommendations to users and removal services are focused on making the removal process easier to follow through and more transparent.

Acknowledgments

We would like to thank all our participants as well as people who helped recruit them, as well as our anonymous reviewers and our shepherd for the suggestions and for helping us improve this paper. This research was supported by the National Science Foundation (grant numbers 2016061, 2016058) and the European Union's Horizon 2020 research and innovation programme(grant agreement no 952226). Our lab has also received gifts from Google.

References

[1] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. 2017. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Comput. Surv.* 50, 3, Article 44 (aug 2017), 41 pages. https://doi.org/10.1145/3054926

[2] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2020. Secrets and likes: the drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology* 30, 4 (2020), 736–758.

[3] Max Aliapoulios, Kejsi Take, Prashanth Ramakrishna, Daniel Borkan, Beth Goldberg, Jeffrey Sorensen, Anna Turner, Rachel Greenstadt, Tobias Lauinger, and Damon McCoy. 2021. *A Large-Scale Characterization of Online Incitements to Harassment across Platforms*. Association for Computing Machinery, New York, NY, USA, 621–638. https://doi.org/10.1145/3487552.3487852

[4] Daly Barnett. 2020. Doxxing: Tips To Protect Yourself Online & How to Minimize Harm. https://www.eff.org/am/deeplinks/2020/12/doxxing-tips-protect-yourself-online-how-minimize-harm. Accessed: 2022-02-23.

[5] Emma Barnett. 2010. Facebook's Mark Zuckerberg says privacy is no longer a 'social norm'. *The Telegraph* 1, 11 (2010).

[6] Patrick Biernacki and Dan Waldorf. 1981. Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research* 10, 2 (1981), 141–163.

[7] Privacy Bot. 2021. https://privacybot.io/. Accessed: 2021-11-23.

[8] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. *APA handbook of research methods in psychology*.

[9] Intel Techniques by Michael Bazzell. [n. d.]. Data Removal Guide. https://inteltechniques.com/workbook.html. Accessed: 2021-11-20.

[10] Intel Techniques by Michael Bazzell. [n. d.]. My Experience with PrivacyBot. https://inteltechniques.com/blog/2021/05/14/my-experience-with-privacybot/. Accessed: 2021-11-23.

[11] Privacy Rights Clearinghouse. [n. d.]. Data Brokers list. https://privacyrights.org/data-brokers. Accessed: 2021-11-

20.

[12] Federal Trade Commission et al. 2014. Data brokers: A call for transparency and accountability.

[13] DeleteMe. [n. d.]. Remove Personal Info from Google. https://joindeleteme.com/. Accessed: 2021-10-20.

[14] Periwinkle Doerfler, Andrea Forte, Emiliano De Cristofaro, Gianluca Stringhini, Jeremy Blackburn, and Damon McCoy. 2021. "I'm a Professor, Which Isn't Usually a Dangerous Job": Internet-Facilitated Harassment and Its Impact on Researchers. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 341 (oct 2021), 32 pages. https://doi.org/10.1145/3476082

[15] David M. Douglas. 2016. Doxing: A Conceptual Analysis. *Ethics and Inf. Technol.* 18, 3 (sep 2016), 199–210. https://doi.org/10.1007/s10676-016-9406-0

[16] Privacy Duck. [n. d.]. Privacy Duck Services. https://privacyduck.square.site/. Accessed: 2021-10-20.

[17] World Privacy Forum. [n. d.]. Data Brokers Opt Out List. https://www.worldprivacyforum.org/2013/12/data-brokers-opt-out/. Accessed: 2021-11-20.

[18] Kevin Gallagher, Sameer Patil, and Nasir Memon. 2017. New me: understanding expert and non-expert perceptions and usage of the tor anonymity network. In *Proceedings of the Thirteenth USENIX Conference on Usable Privacy and Security*. 385–398.

[19] Luke Gelinas, Robin Pierce, Sabune Winkler, I Glenn Cohen, Holly Fernandez Lynch, and Barbara E Bierer. 2017. Using social media as a research recruitment tool: ethical issues and recommendations. *The American Journal of Bioethics* 17, 3 (2017), 3–14.

[20] Jennifer Barrett Glasgow. 2018. Data Brokers: Should They Be Reviled or Revered? *The Cambridge Handbook of Consumer Privacy* (2018), 25–46.

[21] Yael Grauer. [n. d.]. New Open Source Project Automates Data Deletion Requests by Email. https://digital-lab-wp.consumerreports.org/2021/05/12/new-open-source-project-automates-data-deletion-requests-by-email/. Accessed: 2021-11-23.

[22] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2020. *"It's a Scavenger Hunt": Usability of Websites' Opt-Out and Data Deletion Choices*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376511

[23] Hana Habib, Yixin Zou, Aditi Jannu, Neha Sridhar, Chelse Swoopes, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2019. An Empirical Analysis of Data Deletion and Opt-out Choices on 150 Websites *(SOUPS'19)*. USENIX Association, USA, 387–406.

[24] Adele E Howe, Indrajit Ray, Mark Roberts, Malgorzata Urbanska, and Zinta Byrne. 2012. The psychology of security for the home computer user. In *2012 IEEE Symposium on Security and Privacy*. IEEE, 209–223.

[25] California Legistlative Information. [n. d.]. California Consumer Privacy Act of 2018. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1121. Accessed: 2021-10-20.

[26] Business Insider. [n. d.]. The Cambridge Analytica whistleblower explains how the firm used Facebook data to sway elections. https://www.businessinsider.in/tech/news/the-cambridge-analytica-whistleblower-explains-how-the-firm-used-facebook-data-to-sway-elections/articleshow/71461113.cms. Accessed: 2021-11-20.

[27] Kanary. [n. d.]. Know what the internet knows about you. https://www.thekanary.com/. Accessed: 2021-10-20.

[28] Kristen Kozinski and Neena Kapur. [n. d.]. How to Dox Yourself on the Internet. https://open.nytimes.com/how-to-dox-yourself-on-the-internet-d2892b4c5954. Accessed: 2022-02-23.

[29] Ada Lerner, Helen Yuxun He, Anna Kawakami, Silvia Catherine Zeamer, and Roberto Hoyle. 2020. Privacy and Activism in the Transgender Community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376339

[30] Nathan Mattise. [n. d.]. 8chan user offers to "swat" GamerGate critic, cops sent to an old address. https://arstechnica.com/tech-policy/2015/01/8chan-tries-swatting-gamergate-critic-sends-cops-to-an-old-address/. Accessed: 2022-02-20.

[31] Allison McDonald, Catherine Barwulor, Michelle L. Mazurek, Florian Schaub, and Elissa M. Redmiles. 2021. "It's stressful having all these phones": Investigating Sex Workers' Safety Goals, Risks, and Practices Online. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 375–392. https://www.usenix.org/conference/usenixsecurity21/presentation/mcdonald

[32] Aleecia M McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *Isjlp* 4 (2008), 543.

[33] Moses Namara, Daricia Wilkinson, Kelly Caine, and Bart P Knijnenburg. 2020. Emotional and Practical Considerations Towards the Adoption and Abandonment of VPNs as a Privacy-Enhancing Technology. *Proc. Priv. Enhancing Technol.* 2020, 1 (2020), 83–102.

[34] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[35] Patricia A Norberg, Daniel R Horne, and David A Horne. 2007. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of consumer affairs* 41, 1 (2007), 100–126.

[36] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376321

[37] OneRep. [n. d.]. https://onerep.com/. Accessed: 2021-10-20.

[38] Zoë Quinn. [n. d.]. Zoë Quinn: What Happened After GamerGate Hacked Me. https://time.com/4927076/zoe-quinn-gamergate-doxxing-crash-override-excerpt/. Accessed: 2022-02-20.

[39] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. 2016. How I Learned to Be Secure: A Census-Representative Survey of Security Advice Sources and Behavior. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*. Association for Computing Machinery, New York, NY, USA, 666–677.

https://doi.org/10.1145/2976749.2978307
[40] Washington Post Rene Sanchez. [n. d.]. Abortion Foes' Internet Site on Trial. https://www.washingtonpost.com/wp-srv/national/longterm/abortviolence/stories/website.htm. Accessed: 2022-02-20.
[41] GDPR Resources and Information. [n. d.]. GDPR: Rights of the data subject: Right to erasure (Article 17). https://www.gdpr.org/regulation/article-17.html. Accessed: 2021-10-20.
[42] Iskander Sanchez-Rola, Matteo Dell'Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos. 2019. Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security (Asia CCS '19)*. Association for Computing Machinery, New York, NY, USA, 340–351. https://doi.org/10.1145/3321705.3329806
[43] Lucy Simko, Ada Lerner, Samia Ibtasam, Franziska Roesner, and Tadayoshi Kohno. 2018. Computer security and privacy for refugees in the United States. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 409–423.
[44] Peter Snyder, Periwinkle Doerfler, Chris Kanich, and Damon McCoy. 2017. Fifteen Minutes of Unwanted Fame: Detecting and Characterizing Doxing. In *Proceedings of the 2017 Internet Measurement Conference (IMC '17)*. Association for Computing Machinery, New York, NY, USA, 432–444. https://doi.org/10.1145/3131365.3131385
[45] Daniel J Solove. 2001. Access and aggregation: Public records, privacy and the constitution. *Minn. L. Rev.* 86 (2001), 1137.
[46] Daniel J Solove. 2021. The myth of the privacy paradox. *Geo. Wash. L. Rev.* 89 (2021), 1.
[47] Latanya Sweeney. 2000. Simple demographics often identify people uniquely. *Health (San Francisco)* 671, 2000 (2000), 1–34.
[48] Kurt Thomas, Devdatta Akhawe, Michael Bailey, Dan Boneh, Elie Bursztein, Sunny Consolvo, Nicola Dell, Zakir Durumeric, Patrick Gage Kelley, Deepak Kumar, et al. 2021. Sok: Hate, harassment, and the changing landscape of online abuse. (2021).
[49] National Network to End Domestic Violence. [n. d.]. Voting and Survivor Privacy. https://www.techsafety.org/voter-registration-privacy. Accessed: 2021-11-20.
[50] Giridhari Venkatadri, Athanasios Andreou, Yabing Liu, Alan Mislove, Krishna P. Gummadi, Patrick Loiseau, and Oana Goga. 2018. Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface. In *2018 IEEE Symposium on Security and Privacy (SP)*. 89–107. https://doi.org/10.1109/SP.2018.00014
[51] Giridhari Venkatadri, Piotr Sapiezynski, Elissa M. Redmiles, Alan Mislove, Oana Goga, Michelle Mazurek, and Krishna P. Gummadi. 2019. Auditing Offline Data Brokers via Facebook's Advertising Platform. In *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1920–1930. https://doi.org/10.1145/3308558.3313666
[52] Computer world. [n. d.]. Doxxing defense: Remove your personal info from data brokers. https://www.computerworld.com/article/2849263/doxxing-defense-remove-your-personal-info-from-data-brokers.html. Accessed: 2021-11-20.
[53] Yael Grauer @yaelwrites on Github. [n. d.]. Big Ass Data Broker Opt-Out List. https://github.com/yaelwrites/Big-Ass-Data-Broker-Opt-Out-List. Accessed: 2021-11-20.
[54] Yixin Zou, Abraham H. Mhaidli, Austin McCall, and Florian Schaub. 2018. "I've Got Nothing to Lose": Consumers' Risk Perceptions and Protective Actions after the Equifax Data Breach. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. USENIX Association, Baltimore, MD, 197–216. https://www.usenix.org/conference/soups2018/presentation/zou

Appendix

In this appendix we outline the screening survey and the interview guide. We note that this is a semi-structured interview guide. Interviews were conversational and follow up questions were asked based on the experiences of individual interviewees.

Screening Questionnaire

We invite you to take part in a research study to learn from you about the experience of removing private information from the internet. We want to better understand the difficulties encountered and the best practices.

We expect that the survey takes 1 - 3 minutes to complete and if contacted for an interview, it will take a maximum of 45 minutes. Each interview participant will be compensated with a $25 Amazon gift card. Participation is voluntary and you can quit the survey at any time.

If there is anything about the study or your participation that is unclear or that you do not understand, if you have questions or wish to report a research-related problem, you may contact us via email.

1. **Age:**
   - 18 - 24
   - 25 - 34
   - 35 - 44
   - 45 - 54
   - 55+
   - Prefer not to say

2. **Region:** Where is your home located?
   - North America/Central America
   - South America
   - Europe - European Union (EU)
   - Europe - non EU

- Africa
- Asia
- Australia
- Other [Text Box]
- Prefer not to say

3. **Gender:**
   - Woman
   - Man
   - Non-binary
   - Prefer not to disclose
   - Prefer to self-describe: [Text Box]

4. **Email:** Please enter your email address. We will use this to contact you only for the purpose of this interview.

5. **Education:** What is the highest level of education you have completed?
   - High school degree or equivalent
   - Bachelor's degree or equivalent
   - Master's degree
   - Doctoral degree
   - Other [Text Box]
   - Prefer not to say

6. Have you tried to remove any personal data (i.e. personal information, sensitive audio-visual material, sensitive search results etc.) from the internet?
   - Yes
   - No

7. If yes, from which of the following? (Check all that apply)
   - Social networking sites: Facebook, Twitter, Instagram etc.
   - Blogging Platforms
   - Shopping websites
   - Government Registries / Public Records i.e. (Voting records)
   - Search engine results i.e. Google Search Results
   - Data stored and processed by companies
   - Data brokers/ People Search Websites i.e. Intelius, Acxiom, WhitePages, Spokeo
   - Other: [Text Box]

8. If yes, was the information publicly accessible using the internet?
   - Yes
   - No

9. If yes, have you tried to remove data on the basis of any of the following:
   - General Data Protection Regulation (GDPR)
   - California Consumer Privacy Act (CCPA)
   - No
   - Unsure

10. Which of the following actions have you taken (Check all that apply):
    - Paid services such as Privacy Duck, DeleteMe, One Rep etc.
    - Hired a privacy consultant
    - Communicated with a public official (example county clerk)
    - I did it myself
    - Other: [Text Box]

Interview Guide

Thank you for taking the time to participate in this interview.

1. Could you tell us, what do you do for a living? What does that entail?
2. What kind of things do you like to do online?
3. What concerns, if any, have you had while online? What about privacy concerns?
4. PII (Personal Identifiable Information) is defined as information: that directly identifies an individual (e.g., name, address, social security number or combinations of them). What factors do you consider when disclosing your PII or sharing data online or offline?
5. What measures have you taken to remove your personal information from the internet?
6. What motivated you (to look into removing your personal information from the internet)?
7. What type of data were you interested in removing?
8. Can you think of an example when a person or a third party/someone found or shared your private data or information without your knowledge?
9. If you remember, could you name some people search websites (PSW) you might have encountered while browsing online?
10. From your point of your view, how do People Search Websites work?
11. **If checked the "I did it myself" option on question 10 (pre-screening survey):** Can you tell us about the experience of trying to remove your personal information?

12. **If checked the "Paid services such as Privacy Duck, DeleteMe, One Rep" option on question 10 (pre-screening survey):** Can you tell us about the experience of using this service?
13. In your point of view, what are the potential consequences of having your personal information visible on the PSW?
14. How do you feel about the outcome of the removal attempts from the PSW?
15. **If checked any other option on question 7 (pre-screening survey), apart from People Search Websites:** You mentioned that you have removed personal data from [option], can you tell us a bit more about that?
16. Do you have any recommendations for people who might be interested in removing their personal information from these websites?
17. Do you use any applications designed to protect your privacy? Could you name them? (i.e. E2E encrypted IM applications, onion routing, privacy-oriented web browsers)
18. Do you have any other thoughts that you would like to share? Is there anything else we should have asked?